Statistics for diffusion processes with low and high-frequency observations

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. rer. nat.

im Fach Mathematik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Magister Jakub Stanisław Chorowski

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. habil. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Markus Reiß
2. Prof. Dr. Marc Hoffmann
3. Prof. Dr. Mathias Trabs

Tag der mündlichen Prüfung: 22 September 2016

Abstract

In this thesis, we consider the problem of nonparametric estimation of the diffusion coefficients of a scalar time-homogeneous Itô diffusion process from discrete observations under various sampling assumptions.

In the first part, the low-frequency estimation method proposed by Gobet, Hoffmann and Reiß is modified to cover the case of random sampling times. The estimator is shown to be optimal in the minimax sense and adaptive to the sampling distribution. Moreover, Lepski's method is applied to adapt to the unknown Sobolev smoothness of the drift and volatility coefficients.

In the second part, we address the problem of volatility estimation from equidistant observations without a predefined frequency regime. In the case of a stationary diffusion with compact state space and boundary reflection, we introduce a universal estimator that attains the minimax optimal convergence rates for both low and high-frequency observations. Being based on the spectral method, the low-frequency analysis is similar to the study conducted by Gobet, Hoffmann and Reiß. On the other hand, the derivation of the convergence rates in the high-frequency regime requires local averaging of the low-frequency estimator, which makes it mimic the behaviour of the classical high-frequency estimator introduced by Florens-Zmirou.

The analysis of the universal estimator requires tight upper bounds on the estimation error of the occupation time functional for non-continuous functions. In the third part of the thesis, we thus consider the Riemann sum approximation of the occupation time functional of a stationary, time-reversible Markov process. Upper bounds on the squared mean estimation error are provided. In the case of diffusion processes, convergence rates for Sobolev regular functions are obtained.

Zusammenfassung

Diese Dissertation betrachtet das Problem der nichtparametrischen Schätzung der Diffusionskoeffizienten eines ein-dimensionalen und zeitlich homogenen Itô-Diffusionsprozesses. Dabei werden verschiedene diskrete Sampling Regimes untersucht.

Im ersten Teil zeigen wir, dass eine Variante des von Gobet, Hoffmann und Reiß konstruierten Niedrigfrequenz-Schätzers auch im Fall von zufälligen Beobachtungszeiten verwendet werden kann. Wir beweisen, dass der Schätzer optimal im Minimaxsinn und adaptiv bezüglich der Verteilung der Beobachtungszeiten ist. Außerdam wenden wir die Lepski Methode an um einen Schätzer zu erhalten, der zusätzlich adaptiv bezüglich der Sobolev-Glattheit des Drift- und Volatilitätskoeffizienten ist.

Im zweiten Teil betrachten wir das Problem der Volatilitätsschätzung für äquidistante Beobachtungen. Im Fall eines stationären Prozesses, mit kompaktem Zustandsraum, erhalten wir einen Schätzer, der sowohl bei hochfrequenten als auch bei niedrigfrequenten Beobachtungen die optimale Minimaxrate erreicht. Die Konstruktion des Schätzers beruht auf spektralen Methoden. Im Fall von niedrigfrequenten Beobachtungen ist die Analyse des Schätzers ähnlich wie diejenige in der Arbeit von Gobet, Hoffmann und Reiß. Im hochfrequenten Fall hingegen finden wir die Konvergenzraten durch lokale Mittelwertbildung und stellen daubt eine Verbindung zum Hochfrequenzschätzer von Florens-Zmirou her.

In der Analyse unseres universalen Schätzers benötigen wir scharfe obere Schranken für den Schätzfehler von Funktionalen der *Occupation time* für unstetige Funktionen. Wir untersuchen eine auf Riemannsummen basierende Approximation der *Occupation time* eines stationären, reversiblen Markov-Prozesses und leiten obere Schranken für den quadratischen Fehler her. Im Fall von Diffusionsprozessen erhalten wir Konvergenzraten für Sobolev Funktionen.

Contents

1. Introduction **1**

2. Construction and properties of reflected diffusion processes **17**
2.1. Construction 17
2.2. Path properties 21
2.3. The Markovian nature 27

3. Spectral estimation with random sampling times **31**
3.1. Introduction 31
3.2. Minimax convergence rates 38
3.3. Adaptive estimation 39
3.4. Numerical example 40
3.5. Proofs of the upper bounds 43
3.6. Proof of the lower bounds 54
3.7. Proof for the adaptive estimator 57

4. Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies **61**
4.1. Introduction 61
4.2. Discussion 66
4.3. High-frequency analysis 73
4.4. Low-frequency analysis 101

5. Estimating the occupation time **109**
5.1. Introduction 109
5.2. Preliminaries 110
5.3. Convergence rates 111
5.4. Examples 114

6. Conclusion **119**

A. Stability of the eigenvalue problems **121**
A.1. Compact, positive-definite, self adjoint operator 121
A.2. Bilinear coercive form 122
A.3. Generalized eigenvalue problem for positive definite symmetric matrix pairs . . 127

Bibliography **131**

List of Figures **137**

List of Tables **139**

1. Introduction

In 1828, Brown [17] observed that particles suspended in a fluid execute a continuous jittery motion. Since then the understanding of the diffusion processes became an important goal for both physicists and mathematicians. The first successful explanation of this phenomenon was presented by Einstein [33] in 1905, who derived the diffusion equation from the microscopic movements of particles. Almost two decades later, in 1923, Wiener [95] developed a solution: the Brownian motion. In the meantime, physical applications led to generalizations of Einstein's equation: Smoluchowski [81] introduced a space dependent drift and Chapman [18] the diffusion coefficient, also known as the volatility. The physical interpretation of the drift as an external force and the volatility as the temperature have rendered them fundamental concepts in the theory of diffusion processes.

The Brownian assumption of independent increments is not valid for diffusion models with space varying coefficients. After the appearance of the drift and the volatility, it got superseded by the Markov condition stating that the future evolution of the process depends on the past only by the present. The Markov approach moved the focus from the coefficients to the transition semigroup. The semigroup, together with the initial condition X_0, describes the law of the process uniquely and its kernel functions can be interpreted as transition probabilities.

Drift and volatility became important again with the development of stochastic analysis. They appear in the related stochastic differential equation and describe the martingale and bounded-variation terms in the semimartingale decomposition. The Markov and the stochastic analysis approach complement each other, a good example of which is the problem of estimating the diffusion coefficients. Depending on the structure of available observations, the estimation is based on the semimartingale structure of the process (see for example [38, 48, 49, 9]) or uses methods of the Markov theory (see [46, 44, 23]).

The applications of diffusion processes quickly went beyond their origins in physics. In many fields, they became a natural tool for modeling time-homogeneous Markov processes in continuous time. Diffusions are basic models in the theory of thermodynamics and fluid mechanics. By biologists they are used to describe the spread of diseases or the population dynamics. They play a central role in modern mathematical finance, where they are used to model stock prices, as in the famous Black-Scholes option model, but also to model the dynamics of interest rates, foreign exchange rates, and many others.

1.1. The estimation problem

From the point of view of stochastic analysis, a time homogeneous Itô diffusion is characterized for a drift b and volatility σ, as a solution of the following stochastic differential equation (SDE):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{1.1}$$

where W is a standard Brownian motion.

When proposing a diffusion model, one of the main challenges is the appropriate choice of the coefficients b and σ. In many cases, prior knowledge or intuition about the structure of the model allows specifying a family of possible models, each described by a finite set of scalar parameters. In this case, the problem is to identify the parameters of the observed model. On the other hand, it may happen that there is a need to infer the coefficients without any predefined conditions, for example to test some assumptions concerning the parameters. The former type of problems is called parametric whereas the latter is referred to as nonparametric. Although the lack of a priori assumptions is at the core of the nonparametric estimation, it often turns out that the performance of estimation methods depends on certain properties of targeted objects. Consequently, it is convenient to analyze nonparametric estimators under regularity conditions.

The problem of estimating coefficients of a time homogeneous diffusion process depends essentially on the structure of the available observations. Assume that we observe a path of the diffusion process X at equidistant times $0, \Delta, 2\Delta, ..., N\Delta$. A typical assertion for the analysis of estimation procedures is that the sample size $N + 1$ grows to infinity. This can be satisfied by either letting the time horizon of the observations $T = N\Delta$ grow to infinity or by decreasing the time Δ between observations. When the sampling distance Δ is fixed the observations form a lengthening Markov chain that provides information about long-time behaviour of the process. Such observations are called low-frequency. The other possibility are high-frequency observations, when $\Delta = T/N$ for some fixed $T > 0$ decreases to 0. In that case, the discrete observations provide better and better insight into the underlying continuous path $(X_t, 0 \leq t \leq T)$.

In this thesis, we contribute to the problem of the nonparametric estimation of the diffusion coefficients in various aspects. The main result is the construction of the first known universal volatility estimator that attains the optimal minimax convergence rates under both high and low-frequency assumptions, see Chapter 4 for more details. In Chapter 3, we address the problem of nonparametric estimation of both b and σ when the times between consecutive observations

$$\Delta_n = \tau_n - \tau_{n-1}, \quad n = 1, ..., N$$

are identically distributed independent random variables, independent of the diffusion process X.

While analyzing the performance of the universal estimator, we faced a problem of estimating the values of the occupation time functional $\Gamma_T(f) = \int_0^T f(X_s)ds$ for non-continuous functions f, which is an important problem of independent interest. In Chapter 5, we present upper bounds on the convergence rate of the Riemann type estimator $\widehat{\Gamma}_T(f) = \frac{1}{N}\sum_{n=0}^{N-1} f(X_{n\Delta})$. A more detailed outline of the thesis can be found in Section 1.3.

1.2. Literature review

The difficulty of the estimation of the diffusion coefficients depends essentially on the type of observations available. Under the assumption of continuous observations, the volatility is perfectly identifiable via the quadratic variation process. The problem of parametric drift estimation with continuous data was first addressed by Brown and Hewitt [16], where the authors showed that the Maximum Likelihood Estimator is asymptotically normal. A bit later, Banon [10] proposed a nonparametric recursive drift estimator. The estimation was based on the observation that when the diffusion process is stationary, the drift coefficient can be deduced from the derivative of the invariant density. Banon's work started a long line of research, see Bosq [15] for further references.

Assuming continuous observations simplifies the estimation problem but is obviously unrealistic. No matter how the data is measured, in practice one can collect only a finite number of observations. Hence, it is necessary to consider various discretization schemes. A natural relaxation of the continuous data assumption is to assume high-frequency sampling, so that the observations form a good approximation of the continuous path. It must be stressed that considering discrete observations complicates the estimation problem considerably. In general, the transition probabilities of diffusions can not be expressed explicitly, hence the likelihood based estimation is not feasible.

The first parametric estimator of both drift and diffusion coefficients of a discretely observed stationary diffusion was presented in Dacunha-Castelle and Florens-Zmirou [27], c.f. Dohnal [30]. The estimation method was based on small time expansions of the transition densities. In subsequent years other methods were developed such as martingale estimating functions [14], indirect inference [45], approximate maximum likelihood [68, 3, 70] and Bayesian analysis [34, 66].

The nonparametric estimation was first addressed in Florens-Zmirou [38]. The constructed estimator was proved to be consistent and asymptotically normal; Hoffmann [49] argued that it is minimax optimal in the class of diffusion processes with Lipschitz regular diffusion coefficients. The Florens-Zmirou estimator was further generalized in Bandi and Phillips [9]. A drawback of the Florens-Zmirou method is that the precision of the estimation does not improve when imposing higher smoothness constraints on the volatility. This problem was solved by Hoffmann [47] and independently by Jacod [53] by considering regularizing sequences of stopping times.

The estimation from low-frequency data remained an open problem even after the development of the high-frequency methods. In 1998, Hansen et al. [46] proposed the spectral approach to identify scalar diffusions via the eigenpairs of the underlying transition operator. Soon after, Kessler and Sørensen [56] started to analyze the parametric efficiency of spectral methods, ultimately obtaining $\sqrt{N}-$consistent estimators with precise asymptotic properties, see Sørensen [83]. In the nonparametric setting, Gobet et al. [44] constructed a spectral estimator and proved its minimax optimality. In Chorowski and Trabs [23] we further generalized the spectral method, proving that it is minimax optimal even for observations with random independent sampling times. Moreover, we proposed an estimation procedure adaptive with respect to the regularity of the coefficients. Recently, an alternative methodology was developed that adapts the Bayesian inference to the nonparametric setting, see [66, 69, 91, 65].

1.3. Own contributions and outline of the Thesis

A standard simplifying assumption in the low-frequency analysis is that we observe a diffusion process that lives on a compact state space with reflecting barriers, see [44, 65]. Since the high-frequency analysis requires a good understanding of the path properties, we devote Chapter 2 to the discussion of time-homogeneous Itô diffusion processes with boundary reflection.

In Chapter 3, which is based on [23], we consider a scalar diffusion model observed discretely at random times $0, \tau_1, \tau_2, ..., \tau_N$. We argue that when the observation distances

$$\Delta_n = \tau_n - \tau_{n-1}, \quad n = 1, ..., N$$

form an independent and identically distributed sequence of positive random variables with distribution γ, the transition operator R of the observed Markov chain satisfies

$$R = \int_0^\infty P_t \gamma(dt).$$

Hence, the operator R has the same eigenfunctions as the generator, while its eigenvalues are the eigenvalues of the generator modified by the Laplace transform of the distribution γ. We modify the spectral estimators of b and σ proposed by [44] and show that they remain minimax optimal. Moreover, applying Lepski's method, we propose an estimator that adapts to the unknown Sobolev regularity of the coefficients. The first adaptive estimator based on low-frequency observations of a diffusion process has been constructed only recently in [82].

Almost all existing estimation methods are designed to work in a predefined frequency regime. In applications, it means that for given data, the choice of adequate asymptotics must be based on an a priori assumption of the observation frequency scale. Therefore, an important problem is to develop universal methods, that would perform optimally regardless of the sampling frequency. In the parametric setting, the problem of the universal scale estimation was first raised in Jacobsen [51, 52]. Constructed estimators were consistent and asymptotically Gaussian for all values of Δ, but nearly efficient for only small values of Δ. In Chapter 4, which is based on [21], we construct a nonparametric universal estimator and show that it attains for diffusions with weakly differentiable volatility the minimax optimal rates in both high and low-frequency regimes.

As explained in Section 1.6 below, spectral estimation in the high-frequency regime raises new technical difficulties. One of the problems is that the operator that plays the role of the infinitesimal generator has non-differentiable coefficients. Since such operators are not considered in the existing literature that focuses on operators with smooth coefficients, we develop the perturbation theory for elliptic differential operators with Hölder regular coefficients, see Appendix. Furthermore, we derive bounds on the spectral gap and the extrema of the principal eigenfunctions.

When the time horizon of the observations is fixed the invariant density of the process is not observed. Instead, one needs to consider the occupation density. Corresponding estimation problem, faced while considering the performance of the spectral estimator in the high-frequency regime, is to effectively approximate the occupation time of an interval. Surprisingly, this problem was considered only recently by Ngo and Ogawa [64] and Kohatsu-Higa et al. [58] by means of the stable convergence and Malliavin calculus, respectively. In Chapter 5, we analyze the problem of estimating the occupation time functional of Markov processes. More precisely, when X is a stationary time-reversible Markov process, we derive upper bounds on

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{n=1}^{N} f(X_{(n-1)/N}) - \int_0^1 f(X_r)dr\right)^2\right]^{1/2}$$

that depend on the action of fractional operators $(I-L)^{s/2}$, $0 < s < 1$ where L is the infinitesimal generator of X, on f. In the case of X being a scalar diffusion on $[0,1]$ with reflecting barriers, we imply convergence rate $n^{-\frac{1+s}{2}}$ for functions belonging to fractional Sobolev spaces of order s.

In the remaining part of the introduction we present an overview of the existing methods of the nonparametric estimation of the diffusion coefficients in the high and low-frequency regimes. Furthermore, we discuss at an intuitive level the spectral estimation method applied to the high-frequency data. We motivate that the spectral estimators can perform well in the high-frequency regime but highlight the arising technical difficulties.

1.4. Link to nonparametric regression

Intuitively, $b(x)$ and $\sigma(x)$ can be interpreted as the instantaneous conditional mean and variance of the process X, when $X_t = x$. More precisely, we have the identities

$$\lim_{t\to 0} \frac{\mathbb{E}\left[X_t - x | X_0 = x\right]}{t} = b(x) \tag{1.2}$$
$$\lim_{t\to 0} \frac{\mathbb{E}\left[(X_t - x)^2 | X_0 = x\right]}{t} = \sigma^2(x). \tag{1.3}$$

Assume that we are given equidistant observations $X_0, X_\Delta, X_{2\Delta}, ..., X_{N\Delta}$ of a time-homogeneous diffusion process X. Let for simplicity $b = 0$ in equation (1.1) and consider the squared increment:

$$Y_{n\Delta} = \frac{1}{\Delta}(X_{(n+1)\Delta} - X_{n\Delta})^2 = \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds + \epsilon_n,$$

where

$$\epsilon_n = \frac{1}{\Delta}\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma(X_s)dW_s\Big)^2 - \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds.$$

ϵ_n are centered and uncorrelated random variables. Assuming that the time Δ between two consecutive observations tends to zero, we can identify the quantity $\sigma^2(X_s)$ from

$$\frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds,$$

provided that σ is sufficiently regular. Hence, the problem of estimating the volatility can be transferred to a nonparametric regression setting with:

$$Y_{n\Delta} \simeq \sigma^2(X_{n\Delta}) + \epsilon_n, \quad n = 0, ..., N.$$

In [38], Florens-Zmirou introduced the Nadaraya-Watson estimator of σ^2 with a uniform kernel, namely:

$$\widehat{\sigma}^2_{FZ}(x) = \frac{\sum_{n=1}^N \mathbf{1}(|X_{n\Delta} - x| < h_\Delta)(X_{(n+1)\Delta} - X_{n\Delta})^2}{\Delta\sum_{n=1}^N \mathbf{1}(|X_{n\Delta} - x| < h_\Delta)}.$$

h_Δ is a bandwidth parameter which localizes the estimation. Under some regularity assumptions on the coefficients, the estimator $\widehat{\sigma}^2_{FZ}$ was proved to be consistent, provided the sequence h_Δ is such that $\Delta^{-1}h_\Delta \to \infty$ and $\Delta^{-1}h^4_\Delta \to 0$. Further, if $\Delta^{-1}h^3_\Delta \to 0$ then $\widehat{\sigma}^2_{FZ}$ is asymptotically mixed normal, with the limiting variance depending on the local time of the process X. The minimax properties were analyzed by Hoffmann [49], who proved that under Lipschitz condition for σ the estimator attains the optimal rate $\Delta^{1/3}$. The Florens-Zmirou estimator was further generalized in Bandi and Phillips [9].

As explained by Hoffmann [48], the problem of estimating the volatility differs from the nonparametric regression in many delicate but important aspects. Firstly, the observations $X_{n\Delta}$ are neither independent nor identically distributed. Secondly, the noise terms ϵ_n depend on the observations and are not identically distributed. Finally, the domain where the volatility can be identified is random itself. Indeed, in order to estimate $\sigma^2(x)$ we require that the path $(X_t, t \leq N\Delta)$ visits x. Hence, the values of the volatility function can be inferred only on the interval $[\min_{0\leq t\leq T} X_t, \max_{0\leq t\leq T} X_t]$. The density of the observations is given by the local time of the process X, whose regularity is not linked to the smoothness of the coefficients. This is the root cause of a major drawback of the Florens-Zmirou approach - the performance of the estimator does not increase with higher smoothness of the coefficients. This problem was solved by Hoffmann [48], and independently by Jacod [53], by introducing a sequence of stopping times that regularize the density of the observed data.

1.5. Drift estimation

An immediate conclusion of the previous section is that the volatility coefficient is perfectly identifiable from continuous observations $(X_t, 0 \leq t \leq T)$. This is not the case for the drift coefficient. Indeed, the Girsanov theorem implies that it can not be inferred when the time horizon T is fixed. Therefore, to infer about the drift we need to assume that we observe the path $(X_t, 0 \leq t \leq T)$ at equidistant times $0, \Delta, 2\Delta, ..., N\Delta$ with moderate frequency. More precisely, we require that $\Delta \to 0$ and $N\Delta = T \to \infty$. Motivated by (1.2), we define:

$$\widehat{b} = \frac{\sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h)(X_{(n+1)\Delta} - X_{n\Delta})}{\Delta \sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h)}.$$

Decomposing the increment $X_{(n+1)\Delta} - X_{n\Delta}$ into the drift and martingale parts we obtain that

$$\widehat{b} = \frac{\sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h) \int_{n\Delta}^{(n+1)\Delta} b(X_t)dt}{\Delta \sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h)} + \frac{\sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h) \int_{n\Delta}^{(n+1)\Delta} \sigma(X_t)dW_t}{\Delta \sum_{n=1}^{N} \mathbf{1}(|X_{n\Delta} - x| \leq h)}.$$

Assume for simplicity that X is strictly stationary. The first term concentrates around $b(x)$ with an error that depends only on the smoothness of b. Assuming uniform bounds on the volatility and on the stationary density of X, one can show that the second term is of the order $(Th)^{-1/2}$. Hence, choosing the bandwidth h in a way that balances the two errors we obtain a nonparametric estimator of b.

The estimator $\widehat{b}$ was analyzed by Tuan [90] who proved that it is weakly consistent and asymptotically mixed normal. Similarly to the volatility estimation, the above approach can be considerably refined by using different kernel functions, see for example Bandi and Phillips [9], in which the stationarity condition on X is lifted and the asymptotic distribution of the considered estimator is identified.

1.6. Spectral estimation method

In previous considerations we made the restrictive assumption of high sampling frequency. However, in many situations, sampling at arbitrarily small time intervals is not feasible. Hence, it is also desirable to develop estimation methods that work under the assumption of fixed sampling frequency.

Fix $\Delta > 0$ and consider the diffusion X observed at equidistant times $0, \Delta, 2\Delta, ..., N\Delta$. We want to infer the dynamics of the process under the low-frequency assumption $T = N\Delta \to \infty$. It is important to understand that when the sampling frequency is fixed, the observed sample can not be seen as an approximation of the underlying continuous path. In particular, the estimation can not be based on the interpretation of the coefficients as the infinitesimal mean and variance of the process, which was the core idea in the high-frequency setting. On the other hand, the observations

$$X_0, X_\Delta, X_{2\Delta}, ..., X_{N\Delta}$$

form a lengthening Markov chain. Assuming that the process X is recurrent, we can hope to identify its transition operator and use it to describe the underlying dynamics. This idea was made precise in Hansen et al. [46] and later used by Gobet, Hoffmann and Reiß [44] to define and analyze a low-frequency estimator of both drift and volatility coefficients.

When Δ is fixed, it is very convenient to analyze the diffusion X from the perspective of the Markov semigroup theory.

1.6.1. Semigroup theory of a scalar stationary diffusion process

Assume that the diffusion coefficients b, σ satisfy the local growth condition

$$\exists K > 0 \, \forall x \in \mathbb{R} \quad |b(x)| + |\sigma(x)| \leq K(1 + |x|),$$

and that the volatility is not degenerated with locally integrable squared inverse

$$\forall x \in \mathbb{R} \quad \sigma(x) > 0 \quad \text{and} \quad \sigma^{-2} \in L^1_{loc}(\mathbb{R}).$$

Then, for every initial condition X_0 stochastic differential equation (1.1) has a unique in the sense of the probability law non-exploding weak solution, see [54, Chapter 5, Theorem 5.15 and Remark 5.19]. Assume in addition that the function

$$m(x) = \sigma^{-2}(x) \exp\Big(\int_0^x \frac{2b(y)}{\sigma^2(y)} dy \Big), \quad x \in \mathbb{R}$$

is integrable. Then, SDE (1.1) has a stationary solution (see [46, remark before Assumption 2]) with density

$$\mu(x) = \frac{m(x)}{\int_{-\infty}^{\infty} m(x) dx}. \tag{1.4}$$

Denote by $L^2(\mu)$ the Hilbert space of functions square $\mu-$integrable, equipped with the scalar product

$$\langle f, g \rangle_\mu = \int_{\mathbb{R}} f(x) g(x) \mu(x) dx.$$

For $t \geq 0$ we define the conditional expectation operator $P_t : L^2(\mu) \to L^2(\mu)$ by

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x].$$

Family $(P_t)_{t\geq 0}$ forms a strongly continuous semigroup of contractions, more precisely, it satisfies

- $P_t \circ P_s = P_{t+s}$ holds for all $t, s \geq 0$,
- for any $t \geq 0$ it holds $\|P_t\| \leq 1$,
- for each $f \in L^2(\mu)$ it holds $\lim_{t\to 0} P_t f = f$, where the limit is taken in the $L^2(\mu)$ sense.

The semigroup $(P_t)_{t\geq 0}$ is related to the infinitesimal generator L, defined by:

$$\mathrm{dom}(L) = \left\{ f \in L^2(\mu) : \lim_{t\to 0} \frac{(P_t - I)f}{t} \quad \text{exists} \right\},$$
$$Lf = \lim_{t\to 0} \frac{(P_t - I)f}{t}, \quad \text{for } f \in \mathrm{dom}(L).$$

In the formalism of the functional calculus, operator L generates the semigroup (P_t) by the relation

$$P_t = \exp(tL). \tag{1.5}$$

Furthermore, L is the fundamental link between the transition semigroup and the drift and volatility coefficients of the process. Indeed, using Itô's formula one can easily check that twice differentiable functions with compact support are included in the domain of the generator, and that for such function f it holds

$$Lf(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x).$$

Using the explicit formula (1.4) for the invariant measure, we can express L in the so-called divergence form:

$$Lf(x) = \frac{(f'(x)\sigma^2(x)\mu(x))'}{2\mu(x)}. \tag{1.6}$$

The domain of the generator L can be determined explicitly. Hansen et al. [46, Section 3.3] proved that

$$\mathrm{dom}(L) = \Big\{ f \in L^2(\mu) : f' \text{ exists and is absolutely continuous with}$$
$$\lim_{x\to-\infty} f'(x) = \lim_{x\to+\infty} f'(x) = 0 \text{ and } f'' \in L^2(\mu) \Big\}.$$

For $f, g \in \mathrm{dom}(L)$ integration by parts yields

$$\langle Lf, g\rangle_\mu = -\frac{1}{2}\int_{\mathbb{R}} f'(x)g'(x)\sigma^2(x)\mu(x)dx. \tag{1.7}$$

Hence, L is an unbounded self-adjoint operator on the Hilbert space $L^2(\mu)$. Its spectrum is contained in the negative part of the real axis. We are interested in the eigenpairs of L, more precisely in the solutions of the following.

Eigenproblem 1.1. Find $(\lambda, u) \in \mathbb{R} \times L^2(\mu)$ with $u \neq 0$ such that

$$Lu = \lambda u.$$

It is easy to check that when u is constant we have $Lu = 0$, hence the above problem has always at least one trivial solution $\lambda = 0$ and $u \equiv 1$. While in general it may happen that this is the only solution, Hansen et al. [46, in Section 4.2] identified the sufficient condition that ensures the existence of infinitely many eigenpairs.

Theorem 1.2 ([46, Section 4.2], c.f. [72, Section 2])**.** Assume that σ is continuously differentiable with

$$\lim_{|x|\to\infty} \Big(\sigma'(x) - \frac{2b(x)}{\sigma(x)}\Big)^2 = \infty.$$

Then the generator L has a purely discrete spectrum $\{\lambda_i\}_{i\geq 0}$ consisting of the eigenvalues of the Eigenproblem 1.1. The normalized eigenfunctions $\{u_i\}_{i\geq 0}$ form an orthogonal basis of $L^2(\mu)$, which implies that for any $f \in L^2(\mu)$ we have the representation

$$f = \sum_{i\geq 0} \langle f, u_i\rangle_\mu u_i.$$

Furthermore, when $f \in \mathrm{dom}(L)$ it holds

$$Lf = \sum_{i\geq 0} \langle f, u_i\rangle_\mu \lambda_i u_i$$

as well as

$$\mathrm{dom}(L) = \left\{ f \in L^2(\mu) \,:\, \sum_{i\geq 0} \langle f, u_i\rangle_\mu^2 \lambda_i^2 < \infty \right\}.$$

Under the assumptions of Theorem 1.2 the formula (1.5) can be formally understood as

$$P_t f = \sum_{i\geq 0} \langle f, u_i\rangle_\mu e^{t\lambda_i} u_i.$$

In particular,

$$P_t u_i = e^{t\lambda_i} u_i.$$

Hence, $(e^{t\lambda_i}, u_i)_{i\geq 0}$ are the eigenpairs of the transition operator P_t.

1.6.2. Back to the estimation problem

The core idea of the spectral estimation method is that the volatility and drift coefficients can be expressed in terms of the invariant density and an eigenpair of the infinitesimal generator. Indeed, let (λ, u) be a solution of the Eigenproblem 1.1. By (1.7), for any smooth function f it holds that

$$-\frac{1}{2}\int_{\mathbb{R}} f'(x)u'(x)\sigma^2(x)\mu(x)dx = \langle f, Lu\rangle_\mu = \lambda\langle f, u\rangle_\mu = -\lambda\int_{\mathbb{R}} f'(x)\Big(\int_0^x u(y)\mu(x)dy\Big)dx.$$

Since the set of smooth functions is dense in $L^2(\mu)$, we conclude that

$$\sigma^2(x) = \frac{2\lambda \int_0^x u(y)\mu(y)dy}{u'(x)\mu(x)}. \tag{1.8}$$

We can find a similar representation for the drift function. Since

$$b(x) = \frac{(\mu(x)\sigma^2(x))'}{2\mu(x)}, \tag{1.9}$$

plugging (1.8) in (1.9) we obtain

$$b(x) = \frac{\lambda u(x)}{u'(x)} - \frac{\lambda u''(x) \int_0^x u(y)\mu(x)dy}{\mu(x)(u'(x))^2}. \tag{1.10}$$

Inspired by the idea of the spectral method, Gobet, Hoffmann and Reiß [44] constructed nonparametric estimators of the diffusion coefficients based on equations (1.8) and (1.10). They assumed that the observed process X is a scalar diffusion on $[0,1]$ with instantaneous reflection at the boundaries. The assumption of a compact state space simplifies considerably the analysis. Indeed, it ensures that for T large enough the observations cover the entire state space, that spaces $L^2(\mu)$ are isomorphic, and finally that the eigenfunctions are bounded. The problem of the estimation on the entire real line is discussed in Reiß [72].

Construction of the spectral estimator

Following Gobet et al. [44], we will construct estimators of the invariant density μ and the eigenpair (λ, u), then use the representations (1.8) and (1.10) to define plug-in estimators of the drift and volatility coefficients.

Under the assumption of a discrete spectrum of the infinitesimal generator, the diffusion X is ergodic. For a function $f \in L^2(\mu)$ we define:

$$\langle f, \widehat{\mu}_N \rangle = \frac{1}{N} \sum_{n=0}^{N} f(X_{n\Delta}).$$

Mixing property of X implies that

$$\lim_{N\to\infty} \mathbb{E}\left[|\langle f, \widehat{\mu}_N \rangle - \langle f, \mu \rangle|^2 \right] = 0.$$

Consider an L^2-orthogonal basis $(\psi_j)_{j\geq 1}$ and define the finite dimensional approximation spaces

$$V_J = \text{span}\{\psi_j \,:\, j \leq J\}.$$

Since

$$\mu = \sum_{j\geq 1} \langle \mu, \psi_j \rangle \psi_j,$$

it is natural to estimate the invariant density by

$$\widehat{\mu}_N = \sum_{j\leq J} \langle \psi_j, \widehat{\mu}_N \rangle \psi_j. \tag{1.11}$$

On one hand we want J to be large, so that the truncation error is small. On the other hand, every term $\langle \psi_j, \widehat{\mu}_N \rangle$ is estimated with some precision, hence the total error of the sum on the right-hand side of (1.11) grows with J. The optimal choice of the dimension J is a typical problem of nonparametric estimation, known as the bias-variance trade off.

In the low-frequency regime the infinitesimal generator L is not directly observable. To overcome that difficulty, we will estimate an eigenpair (κ, u) of the transition operator P_Δ and use relation (1.5) to infer that $\lambda = \exp(\Delta\kappa)$. Note first that the eigenpair (κ, u) is the solution of the following

Eigenproblem 1.3. Find $(\kappa, u) \in \mathbb{R} \times L^2(\mu)$ with $u \neq 0$ such that

$$\langle P_\Delta u, v \rangle = \kappa \langle u, v \rangle_\mu, \qquad \text{for every } v \in L^2(\mu).$$

In order to approximate the inner products $\langle P_\Delta u, v \rangle$ and $\langle u, v \rangle_\mu$ we define the symmetric matrices

$$\widehat{P}_{i,j} = \frac{1}{2N} \sum_{n=1}^{N} \psi_i(X_{n\Delta})\psi_j(X_{(n+1)\Delta}) + \psi_j(X_{n\Delta})\psi_i(X_{(n+1)\Delta}),$$

$$\widehat{G}_{i,j} = \frac{1}{2N}\psi_i(X_0)\psi_j(X_0) + \frac{1}{N}\sum_{n=1}^{N-1} \psi_i(X_{n\Delta})\psi_j(X_{n\Delta}) + \frac{1}{2N}\psi_i(X_{N\Delta})\psi_j(X_{N\Delta}).$$

By the mixing property, $\widehat{P}_{i,j}$ and $\widehat{G}_{i,j}$ will converge in the mean to $\langle \psi_i, P_\Delta \psi_j \rangle_\mu = \langle P_\Delta \psi_i, \psi_j \rangle_\mu$ and $\langle \psi_i, \psi_j \rangle_\mu$ respectively when the size of the sample $N \to \infty$. Consequently, the data driven counterpart of the Eigenproblem 1.3 is

Eigenproblem 1.4. Find $(\widehat{\kappa}, \widehat{u}) \in \mathbb{R} \times V_J$ with $\widehat{u} \neq 0$ such that

$$\widehat{P}\widehat{u} = \widehat{\kappa}\widehat{G}\widehat{u}.$$

Let $(\widehat{\kappa}, \widehat{u})$ be the solution of the Eigenproblem 1.4 corresponding to the largest nontrivial eigenvalue. We define the spectral estimators of the drift and volatility coefficients by

$$\widehat{\sigma}^2 = \frac{2\widehat{\lambda}\int_0^x \widehat{u}(y)\widehat{\mu}_N(y)dy}{\widehat{u}'(x)\widehat{\mu}_N(x)},$$

$$\widehat{b} = \frac{(\widehat{\mu}_N(x)\widehat{\sigma}^2(x))'}{2\widehat{\mu}_N(x)},$$

with

$$\widehat{\lambda} = \frac{\log(\widehat{\kappa})}{\Delta}.$$

1.6.3. Spectral estimator applied to the high-frequency data

The spectral estimation method was designed to infer the dynamics of a diffusion under the assumption of fixed sampling frequency. As explained in the previous section, the estimation of the transition operator and of the invariant density rely on ergodicity of the underlying diffusion. Furthermore, the estimation error of the eigenfunction $\widehat{u}$ depends essentially on the so-called spectral gap of the transition operator P_Δ, which is the distance between the largest consecutive eigenvalues. Now, when $\Delta \to 0$ all eigenvalues collapse to 1, hence the spectral gap disappears. At first glance, it seems that the spectral method can not provide any reliable information in the high-frequency regime. As we will see next, this intuition is not entirely true.

Assume that we are given discrete observations $X_0, X_\Delta, ..., X_{N\Delta}$ of the process X with Δ small and $N\Delta = T$ fixed. We construct the spectral estimator as we would do for the low-frequency data. Girsanov's theorem implies that we cannot identify the drift coefficient. Similarly, it is not possible to infer the stationary density. Nevertheless, when Δ tends to 0, the estimator $\widehat{\mu}_N$ converges to the occupation density μ_T, which is defined by the relation

$$\frac{1}{T}\int_0^T f(X_s)ds = \int f(x)\mu_T(x)dx, \tag{1.12}$$

where f is any bounded, measurable function. Recall the estimation of the eigenpair (λ, u) of the infinitesimal generator. When $\Delta \to 0$, matrix $\widehat{P}$ becomes close to $\widehat{G}$ and the Eigenproblem 1.4 does not seem to be informative. However, it is the speed of the convergence $P_\Delta \to I$ that carries the information about the infinitesimal generator. Motivated by the relation

$$L = \lim_{\Delta \to 0} \frac{P_\Delta - I}{\Delta},$$

we define the matrix

$$\widehat{L} = \frac{1}{\Delta}(\widehat{P} - \widehat{G}).$$

Using summation by parts, we obtain that

$$\widehat{L}_{i,j} = -\frac{1}{2N}\sum_{n=0}^{N-1}(\psi_i(X_{(n+1)\Delta}) - \psi_i(X_{n\Delta}))(\psi_j(X_{(n+1)\Delta}) - \psi_j(X_{n\Delta})).$$

When $\Delta \to 0$, the occupation formula (1.12) implies that

$$\widehat{L}_{i,j} \xrightarrow{\mathbb{P}} -\frac{1}{2}\int_0^T \psi_i'(X_s)\psi_j'(X_s)\sigma^2(X_s)ds = -\frac{1}{2}\int \psi_i'(x)\psi_j'(x)\sigma^2(x)\mu_T(x)dx.$$

Hence, $\widehat{L}$ approximates the action of an elliptic operator L_T on V_J, where

$$L_T = \frac{(f'(x)\sigma^2(x)\mu_T(x))'}{2\mu_T(x)}$$

has the same form as the infinitesimal generator (c.f. (1.6)), with the difference that the stationary density is replaced by the occupation density. Most importantly, the representation formula

$$\sigma^2(x) = \frac{2\lambda_T \int_0^x u_T(y)\mu_T(y)dy}{u_T'(x)\mu_T(x)},$$

where (λ_T, u_T) is an eigenpair of L_T, holds. We conclude that the core idea of the spectral estimation method is still valid. Consequently, we can expect that the spectral estimator of the volatility function is at least consistent.

We want to stress the particular difficulties that arise in the high-frequency setting. First of all, the occupation density is random. Secondly, it inherits the almost $1/2-$Hölder regularity of the local time, in particular it is not differentiable. Consequently, the domain of the operator L_T is random and the eigenfunctions are only once differentiable. Low regularity of the occupation density makes its estimation ineffective, even assuming that the coefficients are smooth. In Chapter 4, we show how these problems can be bypassed by considering appropriate approximation spaces V_J and local averages of the spectral estimator.

1.7. Minimax convergence rates

Faced with a statistical inference problem, it is crucial to assess the performance of the available estimation methods. A convenient measure of of the estimation performance is provided by the rates of convergence. While they allow to quantitatively compare different estimators, they leave out the question of optimality of a particular method. One answer to that issue is provided by the minimax theory, which is a set of techniques for finding the worst case behavior of a procedure.

Let $\mathcal{P}$ be a set of models, i.e. a family of probability measures on some fixed measurable space $(\Omega, \mathcal{F})$. Let θ be a functional on $\mathcal{P}$. Assume that for $P \in \mathcal{P}$ we observe a realization $\omega \in \Omega$ indirectly via the observable $X_n(\omega)$. We want to infer the value $\theta(P)$. Consider an estimation procedure $T(X_n)$ and an associated loss function $L(T(X_n), \theta(P))$.

A sequence (φ_n) is the upper rate of the estimation procedure T if

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \varphi_n^{-1} L(T(X_n), \theta(P)) < \infty.$$

Furthermore, the rate (φ_n) is minimax optimal if

$$\liminf_{n\to\infty} \inf_{T} \sup_{P\in\mathcal{P}} \varphi_n^{-1} L(T(X_n), \theta(P)) > 0,$$

where the infimum is taken over all possible estimation procedures.

Example (Volatility estimation)**.** Fix $C > 1$ and consider the family $\mathcal{P} = \{\mathbb{P}_\sigma\}$ of probability measures associated to the diffusion processes with zero drift and volatility function σ satisfying

$$C^{-1} \leq \sigma(x) \leq C, \quad \text{and} \quad |\sigma(x) - \sigma(y)| \leq C|x - y| \quad \text{for all} \quad x, y \in \mathbb{R}.$$

For $N \in \mathbb{N}$ we observe a single path $(X_t, 0 \leq t \leq 1)$ at equidistant times i/N, with $i = 0, ..., N$. Consider the Florens-Zmirou estimator of the volatility coefficient:

$$\widehat{\sigma}^2_{FZ}(x) = \frac{\sum_{n=1}^{N} \mathbf{1}(|X_{n/N} - x| < N^{-1/3})(X_{(n+1)/N} - X_{n/N})^2}{N^{-1} \sum_{n=1}^{N} \mathbf{1}(|X_{n/N} - x| < N^{-1/3})}.$$

We are interested in the values of the volatility on $[0, 1]$. Since the estimation of $\sigma^2(x)$ is meaningful only if the process X hits the point x before the time t, we have to condition the estimation on the event

$$\mathcal{L}_v = \{\omega \in \Omega : (\forall x \in [0, 1])\, L_1(x) \geq v\},$$

where $L_1(x) = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^1 \mathbf{1}(|X_s - x| \leq \epsilon) ds$ is the chronological local time of the process

X. Consequently, we consider the following conditional loss function:

$$L(\widehat{\sigma}, \sigma) = \mathbb{E}\left[\int_0^1 \left(\widehat{\sigma}^2(x) - \sigma^2(x)\right)^2 dx | \mathcal{L}_v\right].$$

Hoffmann [49, Proposition 2] proved that

$$\limsup_{N\to\infty} \sup_{\mathbb{P}_\sigma \in \mathcal{P}} N^{1/3} L(\widehat{\sigma}_{FZ}, \sigma) < \infty,$$

and

$$\inf_{\widehat{\sigma}} \limsup_{N\to\infty} \sup_{\mathbb{P}_\sigma \in \mathcal{P}} N^{1/3} L(\widehat{\sigma}, \sigma) > 0,$$

where the infimum is taken over all measurable functions of the observations. Henceforth, the Florens-Zmirou estimator is minimax optimal in the class of bounded Lipschitz regular volatility functions.

Acknowledgments

In the first place I want to address special thanks to Markus Reiß for his advice and supervision during my studies. I am deeply grateful for his guidance and support while I was writing this thesis. He always believed in me and gave me the freedom necessary to believe in myself.

I want to thank Marc Hoffmann for motivating me in my hunt for the scale optimal estimator. His passion for mathematics and broad knowledge were very helpful. I am especially grateful for explaining me the subtleties of the local time and pointing me to the Florens-Zmirou estimator.

I am specially indebted to Randolf Altmeyer for his friendship and commitment. Without Randolf's help my stay in Berlin and organization of live during studies would be much more difficult. Mathematical collaboration with Randolf was really a great pleasure; I was always impressed by his motivation and passion.

I am thankful to Mathias Trabs for productive collaboration with many discussions. I am very grateful to Dominik Pająk, Sebastian Holtz, Martin Wahl and Moritz Jirak for reviewing parts of this manuscript. I want to thank all my friends and colleagues from the Humboldt Universität, especially my office mate Todor Bilarev and members of the Statistics research group.

The financial support by a PhD scholarship of the Research Training Group 1845 "Stochastic Analysis with Application in Biology, Finance and Physics" is gratefully acknowledged.

Last, but not least I want to thank my family for their support. I address special thanks to my wife Urszula for her encouragement and support. I am deeply indebted to my father in law for his help and understanding.

2. Construction and properties of reflected diffusion processes

The presented theory prepares the ground for the subsequent analysis of the spectral estimation method. Two construction procedures of a reflected diffusion process are described and used to deduce Brownian like upper bounds on the modulus of continuity of the paths. The local time theory is developed. The transition semigroup is described using the spectral theory and Gaussian upper bounds on the transition kernel are derived.

Diffusions with reflecting barriers have rich applications. In finance and economic literature, reflected diffusions are used to describe currency exchange rate target-zone models, in which the exchange rate is allowed to float within two barriers enforced by the monetary authority, for details refer to [8, 59, 87]. Reflected diffusions also appear as the payoff of the so-called "Russian Options", see Shepp and Shiryaev [77]. Among applications in mathematical biology, we recall models for population dynamics in which the total number of individuals is affected by oppositely acting forces, for example spontaneous growth and immigration on one hand and random harvesting or predation on the other, see [74]. Finally, reflected Brownian motion has been shown to describe queueing models experiencing heavy traffic, see [50, 57].

2.1. Construction

The problem of solving stochastic differential equations with reflecting boundary conditions goes back to the famous works of Skorokhod [78, 79]. During the last fifty years, the identification of sufficient and necessary conditions for existence and uniqueness of solutions remained and important and interesting question, see for example [86, 76, 80]. In the simplest case of a one dimensional continuous diffusion process with one or two reflecting barriers, strong solutions can be constructed by applying the solutions of the deterministic Skorokhod problem. In this chapter, we describe two explicit constructions of weak solutions. Both methods rely on an appropriate modification of an unrestricted diffusion process with the state space being the entire real line. Presented constructions formalize the intuition that a reflected diffusion behaves locally like an unrestricted process. Furthermore, they are a useful tool to transfer well known path properties of diffusion processes with natural boundaries to the reflected setting, see for example the proof of Theorem 2.6.

For a bounded measurable drift $b : [0,1] \to \mathbb{R}$ and continuous volatility $\sigma : [0,1] \to \mathbb{R}_+$, satisfying $\inf_{x\in[0,1]} \sigma(x) > 0$, consider the following Skorokhod type stochastic differential equation:

$$\begin{aligned} dX_t &= b(X_t)dt + \sigma(X_t)dW_t + dK_t, \qquad (2.1)\\ X_t &\in [0,1] \text{ for every } t \geq 0, \end{aligned}$$

where $(W_t, t \geq 0)$ is a standard Brownian motion and $(K_t, t \geq 0)$ is some adapted continuous process with finite variation, starting form 0, and such that for every $t \geq 0$

$$\int_0^t \mathbf{1}_{(0,1)}(X_s) dK_s = 0$$

holds. By the Engelbert-Schmidt theorem, boundedness of the drift coefficient together with the volatility function being continuous and strictly positive ensure that SDE (2.1) has a weak solution, see Rozkosz and Słomiński [76, Thm. 4.1].

2.1.1. Construction by reflection

To motivate the first method consider the following construction of a Brownian motion with one reflecting barrier at zero. Given a standard Brownian motion $(W_t, t \geq 0)$, a path of the reflected process can be obtained by simply taking the absolute value. Indeed, the Itô-Tanaka formula implies that

$$|W_t| = \int_0^t \operatorname{sign}(W_s) dW_s + L_t(0),$$

where $(L_t, t \geq 0)$ is the local time process of W. Furthermore, by the Lévy characterization theorem the martingale $\int_0^t \operatorname{sign}(W_s) dW_s$ is a Brownian motion. Since $(L_t(0), t \geq 0)$ is an adapted continuous process with finite variation and increasing on the set $\{t : W_t = 0\}$, the process $|W|$ solves the Skorokhod equation. The above construction works because of the space homogeneity of W. Indeed, since the local dynamics are the same on the entire line, $|W|$ behaves locally like the Brownian motion, in the same time being restricted to the nonnegative values.

In the following, we generalize the above construction to arbitrary diffusion processes. Note that a very similar method, in case of reflecting barriers at -1 and 1, is described in [43, Chapter I.23]. First, we extend the coefficients b, σ to the whole real line.

Definition 2.1. Define $f : \mathbb{R} \to [0, 1]$ by

$$f(x) = \begin{cases} x - 2n & : 2n \leq x < 2n + 1 \\ 2(n + 1) - x & : 2n + 1 \leq x < 2n + 2 \end{cases}, \text{ for } n \in \mathbb{N}.$$

Function f is almost everywhere differentiable with the left derivative

$$f'(x) = \begin{cases} 1 & : 2n < x \leq 2n + 1 \\ -1 & : 2n + 1 < x \leq 2n + 2 \end{cases}.$$

For $\sigma, b : [0, 1] \to \mathbb{R}$ we define the extended coefficients $\widetilde{\sigma}, \widetilde{b} : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} \widetilde{b}(x) &= f'(x) \cdot b \circ f(x) \\ \widetilde{\sigma}(x) &= \sigma \circ f(x). \end{aligned}$$

Theorem 2.2. Let $(Y_t, t \geq 0)$ be a solution of the SDE

$$\begin{aligned} dY_t &= \widetilde{b}(Y_t) dt + \widetilde{\sigma}(Y_t) dW_t, \qquad (2.2) \\ Y_0 &= X_0, \end{aligned}$$

with the initial condition $X_0 \in [0,1]$ and independent of the driving Brownian motion W. Define

$$X_t = f(Y_t).$$

The process $(X_t, t \geq 0)$ is a weak solution of the SDE (2.1).

Proof. Note first that the existence of a weak solution $(Y_t, t \geq 0)$ of the SDE (2.2) follows from the Engelbert-Schmidt theorem, see for example [54, Theorem 5.4 in Chapter 5]. Process Y is a continuous semimartingale, hence by [73, Chapter VI, Theorem 1.2] it admits a local time process $(L_t^Y, t \geq 0)$. By the Itô-Tanaka formula ([73, Chapter VI, Theorem 1.5]) process X satisfies

$$\begin{aligned} X_t &= x_0 + \int_0^t \widetilde{b}(Y_s) f'(Y_s) ds + \int_0^t \widetilde{\sigma}(Y_s) f'(Y_s) dW_s + \sum_{n \in \mathbb{Z}} L_t^Y(2n) - \sum_{n \in \mathbb{Z}} L_t^Y(2n+1) \\ &= x_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s + K_t, \end{aligned}$$

where $B_t = \int_0^t f'(Y_s) dW_s$ and $K_t = \sum_{n \in \mathbb{Z}} L_t^Y(2n) - \sum_{n \in \mathbb{Z}} L_t^Y(2n+1)$. Note that for any $T > 0$ the path $(Y_t, 0 \leq t \leq T)$ is bounded, hence K is well defined. B is a martingale with quadratic variation process

$$\langle B \rangle_t = \int_0^t (f'(Y_s))^2 ds = t.$$

Hence, Lévy's characterization theorem implies that B is a standard Brownian motion. From the properties of the local time process L_t^Y follows that K is adapted and continuous with finite variation, starting from zero and varying on the set

$$\bigcup_{n \in \mathbb{Z}} \{Y_t = 2n\} \cup \{Y_t = 2n+1\} \subseteq \{X_t \in \{0,1\}\}.$$

We conclude that X together with B solve the SDE (2.1). □

2.1.2. Construction by time change

The intuitive understanding of a reflected process is that in the interior of the state space the reflection is not noticeable, while at the boundary it forces the paths to remain inside the state space. Consider an unrestricted diffusion process with coefficients that, when restricted to $[0,1]$, match these of the reflected one. Then, the parts of the path that are contained inside of the interval $[0,1]$ behave like the paths of the reflected diffusion. The formal difficulty is to define a time change that will glue these paths together.

Consider functions $\widetilde{b}$ and $\widetilde{\sigma}$, defined on the whole real line and such that

$$\widetilde{b} \upharpoonright [0,1] = b \text{ and } \widetilde{\sigma} \upharpoonright [0,1] = \sigma.$$

Furthermore, assume that the SDE

$$\begin{aligned} &dX_t = \widetilde{b}(X_t) dt + \widetilde{\sigma}(X_t) dW_t, \\ &X_0 \in [0,1], \end{aligned}$$

has a solution $(\widetilde{X}_t, t \geq 0)$ such that

$$\lim_{t\to\infty} \int_0^t \mathbb{I}_{(0,1)}(\widetilde{X}_s)ds = \infty.$$

Let $L_t(x)$ be a continuous version of the local time of X (see [73, Chapter VI] for the definition and properties). Consider function

$$f(x) = \begin{cases} 0 & : x < 0 \\ x & : 0 \leq x \leq 1 \\ 1 & : x > 1 \end{cases}.$$

Applying the Itô-Tanaka formula we obtain

$$\begin{aligned} Y_t = f(X_t) &= X_0 + \int_0^t \mathbb{I}_{(0,1)}(X_s)\widetilde{b}(X_s)ds + \int_0^t \mathbb{I}_{(0,1)}(X_s)\widetilde{\sigma}(X_s)dW_s + \frac{1}{2}L_t(0) - \frac{1}{2}L_t(1). \\ &= X_0 + \underbrace{\int_0^t b(Y_s)\mathbb{I}_{(0,1)}(Y_s)ds}_{D_t} + \underbrace{\int_0^t \mathbb{I}_{(0,1)}(Y_s)\sigma(Y_s)dW_s}_{M_t} + \underbrace{\frac{1}{2}L_t(0) - \frac{1}{2}L_t(1)}_{K_t}. \end{aligned}$$

Process D is of finite variation and M is a martingale. Let

$$A(t) = \int_0^t \mathbb{I}_{(0,1)}(X_s)ds,$$

be the amount of time up to t that the process X is contained in $(0,1)$. Let

$$C(t) = \inf\{s > 0 | A(s) > t\}$$

be the right-continuous inverse function of A. C is strictly increasing with jumps $[C(t_-), C(t)]$ corresponding to the constancy levels of the process Y. Finally, define the process

$$Z_t = Y_{C(t)}.$$

Since Y is $C-$continuous process (i.e Y_s =const for $s \in (C(t_-), C(t)$, c.f. [73, Chapter 5, Definition 1.3]) Z is a continuous semimartingale adapted to the time changed filtration $\mathcal{F}_C = (\mathcal{F}_{C_t})_{t\geq 0}$, see [73, Proposition 1.4].

Theorem 2.3. Process $(Z_t, 0 \leq t)$ satisfies

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t + dK_t,$$

where B is an $\mathcal{F}_C$-Brownian motion and K is an $\mathcal{F}_C$-adapted continuous process with finite variation, starting form 0, and varying only on the set $\{Z_t \in \{0,1\}\}$. Hence, Z is a weak solution of the SDE (2.1).

Proof. By definition of the process Z it holds

$$Z_t = x + D_{C(t)} + M_{C(t)} + K_{C(t)}.$$

Substituting $s = C(u)$, we get

$$D_{C(t)} = \int_0^{C(t)} b(Y_s)dA(s) = \int_0^t b(Y_{C(u)})du = \int_0^t b(Z_u)du. \tag{2.3}$$

We will construct an $\mathcal{F}_C-$Brownian motion B such that

$$M_{C(t)} = \int_0^t \sigma(Z_s)dB_s. \tag{2.4}$$

First, note that since for $s \in (C(t_-), C(t))$ we have $\mathbf{1}_{(0,1)}(Y_s) = 0$, the process M is $C-$continuous, hence $M_C = (M_{C(t)}, t \geq 0)$ is a continuous $\mathcal{F}_C-$martingale. Its quadratic variation equals

$$\langle M_C \rangle_t = \langle M \rangle_{C(t)} = \int_0^{C(t)} \sigma^2(Y_s)\mathbb{I}_{(0,1)}(Y_s)ds = \int_0^{C(t)} \sigma^2(Y_s)dA(s) = \int_0^t \sigma^2(Z_s)ds.$$

We define $B_t = \int_0^t \sigma^{-1}(Z_s)dM_{C(s)}$ (note that since σ is strictly positive, its reciprocal is well defined). Process B is a continuous local $\mathcal{F}_C-$martingale with quadratic variation

$$\langle B \rangle_t = \int_0^t \sigma^{-2}(Z_s)d\langle M \rangle_{C(s)} = \int_0^t ds = t.$$

Hence, by the Lévy characterization theorem, B is a Brownian motion . By the associativity of the stochastic integral

$$\int_0^t \sigma(Z_s)dB_s = \int_0^t \sigma(Z_s)\sigma^{-1}(Z_s)dM_{C(s)} = M_{C(s)}.$$

From (2.3) and (2.4) it follows that

$$Z_t = x + \int_0^t b(Z_u)du + \int_0^t \sigma(Z_s)dB_s + K_{C(t)},$$

where B is a Brownian motion and K_C is a non-decreasing process with

$$K_{C(0)} = K_0 = \frac{1}{2}L_0(0) - \frac{1}{2}L_0(1) = 0.$$

K_C increases on the set $\left\{X_{C(t)} = 0\right\} \subseteq \left\{Y_{C(t)} = 0\right\} \subseteq \{Z_t = 0\}$ and decreases on $\left\{X_{C(t)} = 1\right\} \subseteq \{Z_t = 1\}$, hence it is varying only on the set $\{Z_t \in \{0, 1\}\}$. $\square$

2.2. Path properties

Seen locally, a reflected diffusion process behaves like an unrestricted Itô diffusion. This observation, which is the corner stone of the constructions presented in Section 2.1 constitutes the fundamental intuition necessary to derive path properties of a reflected process.

Assumption 2.4. *Fix $0 < \lambda < \Lambda$. From now on we assume that drift $b : [0, 1] \to \mathbb{R}$ is*

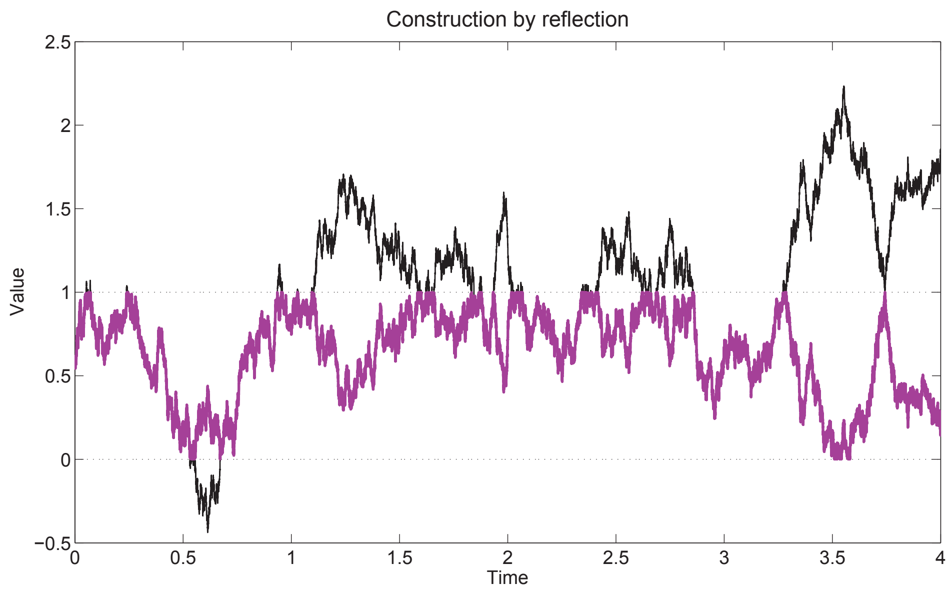

Figure 2.1.: A path of the reflected diffusion process is constructed by reflecting the unrestricted process when crossing the barriers 0 and 1. Note that in order to preserve local dynamics, the coefficients of the unrestricted process are determined by their values in the interval $[0, 1]$.

Figure 2.2.: Given an unrestricted path, the reflected process is constructed by "gluing" together the parts that are contained in the interval $[0, 1]$.

measurable, volatility $\sigma : [0,1] \to \mathbb{R}_+$ *is continuous and that*

$$|b(x)| \leq \Lambda,\ \lambda \leq |\sigma(x)| \leq \Lambda$$

holds for every $x \in [0,1]$.

Throughout the rest of the section we take the Assumption 2.4 as granted. Let X be the solution of the SDE (2.1). Denote by $\mathbb{P}_{\sigma,b}$ the law of the diffusion X on the canonical space Ω of continuous functions over the positive axis with values in $[0,1]$, equipped with the topology of the uniform convergence on compact sets and endowed with its Borel $\sigma-$field $\mathcal{F}$. Denote by $\mathbb{E}_{\sigma,b}$ the corresponding expectation operator.

2.2.1. Modulus of continuity

The modulus of continuity is one of the fundamental tools to measure quantitatively the continuity of functions. In this section, we want to show that the paths of the reflected diffusion X share the same properties as paths of a Brownian motion process.

Definition 2.5. Denote by ω_T the modulus of continuity of the path $(X_t, 0 \leq t \leq T)$, i.e.

$$\omega_T(\delta) = \sup_{\substack{0 \leq s,t \leq T \\ |t-s| \leq \delta}} |X_t - X_s|.$$

Theorem 2.6. Grant Assumption 2.4. For every $p \geq 1$, there exists a positive constant $C_p < \infty$ that depends only on λ and Λ such that

$$\mathbb{E}_{\sigma,b}[\omega_T^p(\Delta)] \leq C_p \Delta^{p/2}(1 \vee \ln(2T/\Delta))^p. \tag{2.5}$$

Proof. Fischer and Nappo [36] proved the above bound for the standard Brownian motion. We will generalize their result to diffusions with boundary reflection.

Step 1. Consider a martingale M satisfying $dM_t = \sigma(X_t)dW_t$. By the Dambis, Dubins-Schwarz theorem, $M_t = B_{\int_0^t \sigma^2(X_u)du}$ for some Brownian motion B. Consequently,

$$|M_t - M_s| = \left|B_{\int_0^t \sigma^2(X_u)du} - B_{\int_0^s \sigma^2(X_u)du}\right| \leq \omega^B\left(\int_s^t \sigma^2(X_u)du\right) \leq \omega^B\left(|t-s|\|\sigma^2\|_\infty\right),$$

where ω^B is the modulus of continuity of B. Thus, (2.5) holds for the martingale M, with a constant that depends only on the uniform upper bound on the squared volatility σ^2.

Step 2. Consider a semimartingale Y satisfying $dY_t = b(Y_t)dt + dM_t$. Then,

$$|Y_t - Y_s| \leq \left|\int_0^t b(Y_u)du - \int_0^s b(Y_u)du\right| + |M_t - M_s| \leq |t-s|\|b\|_\infty + \omega^M(|t-s|).$$

Consequently, (2.5) holds for the semimartingale Y, with a constant that depends only on the upper bounds on σ and b.

Step 3. For (σ, b) satisfying Assumption 2.4 consider the reflected diffusion process X

satisfying the SDE (2.1). Let

$$\begin{aligned} dY_t &= \widetilde{b}(Y_t)dt + \widetilde{\sigma}(Y_t)dW_t, \\ X_t &= f(Y_t), \end{aligned}$$

where $\widetilde{b}, \widetilde{\sigma}$ and f are as in Definition 2.1. From Step 2, it follows that (2.5) holds for the semimartingale Y with a uniform constant that depends only on λ, Λ. By the construction of the reflected process X, together with the inverse triangle inequality, we have $|X_s - X_t| \leq |Y_s - Y_t|$. We conclude that $\omega^X \leq \omega^Y$, hence the claim (2.5) holds for the reflected diffusion X. □

2.2.2. Local time

In this section, we develop the local time theory of reflected diffusion processes. A standard reference for the construction and properties of a semimartingale local time is [73, Chapter VI]. Consider $(X_t, 0 \leq t)$ a solution of the SDE (2.1) with coefficients satisfying Assumption 2.4.

Definition 2.7. Set $t > 0$. For a Borel set $A \subseteq [0,1]$ we define the occupation measure $T_t(A)$ of the path $(X_s, 0 \leq s \leq t)$, with respect to the quadratic variation of X, by

$$T_t(A) = \int_0^t \mathbf{1}_A(X_s)\sigma^2(X_s)ds.$$

When $T_t(A)$ is absolutely continuous with respect to the Lebesgue measure dx on the interval $[0,1]$, we define the local time L by the Radon-Nikodym derivative:

$$L_t(x) = \frac{dT_t}{dx}.$$

Theorem 2.8 (Itô-Tanaka formula)**.** The local time L exists and has a continuous version in both $t > 0$ and $x \in (0,1)$. For every $x \in [0,1]$ the process $(L_t(x), t \geq 0)$ is non-decreasing and increases only when $X_t = x$. Furthermore, if f is the difference of two convex functions, we have

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'_-(X_s)\sigma(X_s)dW_s + \int_0^t f'_-(X_s)b(X_s)ds \\ &+ \frac{1}{2}\int_0^1 L_t(x)f''(dx) + \int_0^t f'_-(X_s)dK_s, \end{aligned} \tag{2.6}$$

where f_- is the left derivative of f.

Proof. Reflected diffusion X is a continuous semimartingale. By [73, Chapter VI, Theorem 1.2 and Theorem 1.5] there exists a process $(L_t(x) : x \in (0,1), t \geq 0)$, continuous and non-decreasing in t, càdlàg in x and such that (2.6) holds. Furthermore, by [73, Chapter VI, Theorem 1.7] for every $x \in (0,1)$

$$L_t(x) - L_t(x_-) = 2\int_0^t \mathbf{1}_{\{X_s=x\}}b(X_s)ds + 2\int_0^t \mathbf{1}_{\{X_s=x\}}dK_s = 0.$$

The concentration of the associated measure dL_t on the set $\{X_t = x\}$ follows from [73, Chapter VI, Proposition 1.3]. $\square$

Lemma 2.9. *For every $T > 0$ and $p \geq 1$ we have*

$$\sup_{(\sigma,b)} \sup_{x\in(0,1)} \mathbb{E}_{\sigma,b}\left[\sup_{t\leq T} L_t^p(x)\right] < \infty,$$

where the first supremum is taken over all coefficients that satisfy Assumption 2.4.

Proof. The usual way to bound the moments of the local time is to use the Itô-Tanaka formula for function $f_x(y) = (y-x)^+$, see e.g. [73, Chapter VI, Theorem 1.7]. Because of the additional reflection term dK_t, we make a less intuitive choice of the function f that guarantees $f'(0) = f'(1) = 0$.

Set $T > 0$, $p \geq 1$ and $x \in (0, 1/2]$. Let

$$f_x(y) = \begin{cases} 3x - 2x^2 & : y < x \\ 3y - 2y^2 & : x \leq y \leq 3/4 \\ 3/8 & : 3/4 < y \end{cases},$$

such that the left derivative of f_x is equal to

$$f'_x(y) = \mathbf{1}(x < y \leq 3/4)(3 - 4y).$$

By the Itô-Tanaka formula (2.6)

$$\begin{aligned} \frac{3-4x}{2} L_t(x) &= f_x(X_t) - f_x(X_0) - \int_0^t \mathbf{1}(x < X_s \leq \tfrac{3}{4})(3 - 4X_s)\sigma(X_s)dW_s \\ &\quad - \int_0^t \mathbf{1}(x < X_s \leq \tfrac{3}{4})(3 - 4X_s)b(X_s)ds + 2\int_0^t \mathbf{1}(x < X_s \leq \tfrac{3}{4})\sigma^2(X_s)ds. \end{aligned}$$

Applying the uniform bounds on b and σ, together with the Burkholder-Davies-Gundy inequality, we conclude that for $t \leq T$ it holds

$$\sup_{x\in(0,1/2)} \mathbb{E}_{\sigma,b}[L_t(x)^p] \leq C_{p,T},$$

with some positive constant $C_{p,T}$ that depends only on λ and Λ. For $x \in (1/2, 1)$ we consider function

$$f_x(y) = \begin{cases} 1/8 & : y < 1/4 \\ y - 2y^2 & : 1/4 \leq y \leq x \\ x - 2x^2 & : x < y \end{cases}$$

and proceed similarly. $\square$

Theorem 2.10. For any $T > 0$, $p \geq 1$ and $x, y \in (0, 1)$ it holds

$$\mathbb{E}_{\sigma,b}\left[\sup_{t\leq T} |L_t(x) - L_t(y)|^{2p}\right] \leq C_{p,T}|x-y|^p, \tag{2.7}$$

where the constant $C_{p,T}$ depends only on the uniform bounds λ, Λ. In particular, the family L of local times can be chosen such that functions $x \longmapsto L_t(x)$ are almost surely Hölder continuous of order α for every $\alpha < 1/2$ and uniformly in $t \leq T$. Moreover, for every $p \geq 1$ and $t \leq T$ it holds

$$\sup_{(\sigma,b)} \mathbb{E}_{\sigma,b}\left[\sup_{x\in[0,1]} L_t^p(x)\right] < \infty, \tag{2.8}$$

where the supremum is taken over all coefficients satisfying Assumption 2.4.

Proof. The proof goes along the same lines as [73, Chapter VI Theorem 1.7]. We will first show the inequality (2.7). For $x \in (0,1)$, by the Itô-Tanaka formula (2.6) we have

$$\begin{aligned}\frac{1}{2}L_t(x) = (X_t - x)^+ - (X_0 - x)^+ - \int_0^t \mathbf{1}(X_s > x)\sigma(X_s)dW_s +\\ - \int_0^t \mathbf{1}(X_s > x)b(X_s)ds - \int_0^t \mathbf{1}(X_s = 1)dK_s.\end{aligned}$$

Since the function $x \longmapsto (X_t - x)^+ - (X_0 - x)^+ - \int_0^t \mathbf{1}(X_s = 1)dK_s$ is uniformly Lipschitz, we need only to consider the martingale term $M_t^x = \int_0^t \mathbf{1}(X_s > x)\sigma(X_s)dW_s$ and the finite variation term $D_t^x = \int_0^t \mathbf{1}(X_s > x)b(X_s)ds$. For $x, y \in (0,1)$, $x < y$ Hölder's inequality together with Lemma 2.9 yield

$$\begin{aligned}\mathbb{E}_{\sigma,b}\Big[\sup_{t\leq T}|D_t^x - D_t^y|^{2p}\Big] &\leq \mathbb{E}_{\sigma,b}\Big[\Big(\int_0^T \mathbf{1}(y > X_s > x)|b(X_s)|ds\Big)^{2p}\Big]\\ &\leq \Lambda^{2p}\,\mathbb{E}_{\sigma,b}\Big[\Big(\int_x^y L_T(z)dz\Big)^{2p}\Big]\\ &\leq \Lambda^{2p}\,|y-x|^{2p-1}\Big|\int_x^y \mathbb{E}_{\sigma,b}[L_T^{2p}(z)]dz\Big| \leq C_{p,T}|y-x|^{2p},\end{aligned}$$

for some constant $C_{p,T} > 0$ that depends only on λ, Λ. To bound the increments of the martingale M^x we use the Burkholder-Davies-Gundy inequality together with Hölder's inequality, obtaining

$$\begin{aligned}\mathbb{E}_{\sigma,b}\Big[\sup_{t\leq T}|M_t^x - M_t^y|^{2p}\Big] &\leq C_p\mathbb{E}_{\sigma,b}\Big[\Big(\int_0^T \mathbf{1}(y > X_s > x)\sigma^2(X_s)ds\Big)^{p}\Big]\\ &\leq C_p\Lambda^{2p}\,\mathbb{E}_{\sigma,b}\Big[\Big(\int_x^y L_T(z)dz\Big)^{p}\Big] \leq \widetilde{C}_{p,T}|y-x|^{p}.\end{aligned}$$

We finished the proof of the bound (2.7). From the Kolmogorov continuity criterion (see [73, Chapter I, Theorem 2.1]) follows that there exists a modification $\widetilde{L}$ of the family of local times, such that functions $x \longmapsto \widetilde{L}_t(x)$ are almost surely Hölder continuous of order α for every $\alpha < 1/2$ and uniformly in $t \leq T$. Furthermore, for any $\alpha < 1/2$ and $p \geq 2$ it holds

$$\sup_{(\sigma,b)} \mathbb{E}_{\sigma,b}\Big[\Big(\sup_{x\neq y}\frac{|\widetilde{L}_t(x) - \widetilde{L}_t(y)|}{|x-y|^{\alpha}}\Big)^{p}\Big] < \infty.$$

Fix $x_0 \in (0,1)$. By the bound above and Lemma 2.9 we conclude that

$$\sup_{(\sigma,b)} \mathbb{E}_{\sigma,b}\Big[\sup_{x\in[0,1]} \widetilde{L}_t^p(x)\Big] \le \sup_{(\sigma,b)} \mathbb{E}_{\sigma,b}\Big[\Big(\sup_{x\neq x_0} \frac{|\widetilde{L}_t(x)-\widetilde{L}_t(x_0)|}{|x-x_0|^\alpha} + \widetilde{L}_t(x_0)\Big)^p\Big] < \infty. \qquad \square$$

2.3. The Markovian nature

In the previous section, we investigated the path properties of a reflected diffusion process. The core observation was that, due to the lower bound on the volatility, the paths of the reflected diffusion have similar regularity as a standard Brownian motion. In this section, we want to analyze the reflected diffusion from the point of view of the Markov theory.

Consider (σ, b) satisfying Assumption 2.4 with σ weakly differentiable. Let X be a weak solution of the SDE (2.1) with some initial condition X_0 independent of the driving Brownian motion W. Define

$$\mu(x) = C_0\sigma^{-2}(x)\exp\Big(\int_0^x 2b(y)\sigma^{-2}(y)dy\Big),$$

where C_0 is a normalizing constant such that μ is a density of a probability measure on $[0,1]$. Furthermore, define the $\mu-$induced scalar product

$$\langle f,g\rangle_\mu = \int_0^1 f(x)g(x)\mu(x)dx$$

and corresponding Hilbert space $L^2(\mu)$. Since under Assumption 2.4 density μ is bounded by positive constants from above and below, the space $L^2(\mu)$ is equivalent to $L^2([0,1])$. For $t>0$ denote by P_t the conditional expectation operator corresponding to X, i.e.

$$P_tf(x) = \mathbb{E}_{\sigma,b}[f(X_t)|X_0 = x]. \tag{2.9}$$

Operators $(P_t, t\ge 0)$ form a strongly continuous semigroup of contractions on $L^2(\mu)$. Using Itô's formula, we can identify its infinitesimal generator L as

$$\begin{aligned}\operatorname{dom}(L) = \{f \in L^2(\mu) \,:\, & f' \text{ exists and is absolutly continuous}\\ & \text{with } f'(0)=f'(1)=0 \text{ and } f''\in L^2(\mu)\},\\ Lf = bf' + \frac{1}{2}&\sigma^2 f''.\end{aligned}$$

Note that the Neumann boundary conditions correspond to the boundary reflection, see [46] for a detailed discussion. Generator L can be expressed in the so-called divergence form

$$Lf(x) = \frac{(\sigma^2(x)\mu(x)f'(x))'}{2\mu(x)}. \tag{2.10}$$

Integration by parts implies that L is a self-adjoint non-positive operator on $L^2(\mu)$. Indeed, for $f,g\in\operatorname{dom}(L)$ it holds

$$\langle f, Lg\rangle_\mu = \frac{1}{2}\int_0^1 f(x)(\sigma^2(x)\mu(x)g'(x))' = -\frac{1}{2}\int_0^1 f'(x)g'(x)\sigma^2(x)\mu(x)dx. \tag{2.11}$$

Furthermore, [19, Proposition 4.17] states that L has a compact resolvent operator. We conclude that L has purely discrete spectrum and its eigenfunctions form an orthogonal basis of $L^2(\mu)$. Denote by (ξ_i, u_i) the eigenpairs of L. By Plancherel's theorem, every $f \in L^2(\mu)$ has a unique representation

$$f = \sum_{i \geq 0} \langle f, u_i \rangle_\mu u_i.$$

The conditional expectation operator is related to the infinitesimal generator via

$$P_t f = \sum_{i \geq 0} \langle f, u_i \rangle_\mu e^{t\xi_i} u_i,$$

which in the sense of the functional calculus can be simply written as $P_t = e^{tL}$.

2.3.1. Off-diagonal Gaussian upper bounds on the transition kernel

The conditional expectation operator P_t defined by (2.9) admits a density function p_t with respect to the Lebesgue measure, i.e. for any $f \in L^2(\mu)$ we have

$$P_t f(x) = \int_0^1 p_t(x, y) f(y) dy.$$

Adapting standard Markovian notation we will call p_t the transition kernel. Note that p_t is the heat kernel of the generator L, as for any $x \in [0, 1]$ the function

$$u(t, y) = p_t(x, y)$$

it is the fundamental solution of the heat equation

$$Lu = \frac{\partial u}{\partial t},$$

with initial condition

$$\lim_{t \to 0} u(t, y) = \delta(x - y),$$

where δ denotes the Dirac delta function. As such, kernel p_t plays a fundamental role in the understanding of the dynamics of the diffusion X and of the spectral properties of the generator L.

When X is a one dimensional diffusion with smooth coefficients the existence of Gaussian upper bounds on the transition kernel follows from the general theory of partial differential equations, see e.g. [39, Chapter 9]. As demonstrated in the Example 2.11 below some regularity of the coefficients is necessary to ensure the existence of the off-diagonal bounds. This regularity condition can be substantially weakened for diffusion processes with generator being an elliptic operator in the divergence form. Although smooth coefficients are usually assumed to simplify the exposition, in case of divergence operators the value of the upper bound depends only on the size of the coefficients. This observation was used in [75] to replace the continuous drift assumption by an integrability condition.

The Gaussian off-diagonal bounds on the transition kernel of a diffusion process with generator being an elliptic divergence operator are well established, see for example [32, 85, 28] and [13, Chapter VII]. More recently, Gaussian upper bounds were derived for multivariate diffusions on manifolds with the infinitesimal generator satisfying Neumann boundary conditions, see [96].

Example 2.11. Fix $y \in \mathbb{R}$ and $\beta > 0$. By [71, Theorem 2] the diffusion process X defined by the SDE

$$dX_t = \beta \operatorname{sign}(y - X_t)dt + dW_t,$$

admits a transition kernel

$$p_t^\beta(x, y) = \frac{1}{\sqrt{2\pi t}} \int_{\frac{|x-y|}{\sqrt{t}}}^{\infty} z \exp\Big(- \frac{(z - \beta\sqrt{t})^2}{2} \Big) dz.$$

We will show that for p_t^β the Gaussian upper bound does not hold. Assume by contradiction that for some $c_1, c_2 > 0$ we have

$$p_t^\beta(x, y) \leq \frac{c_1}{\sqrt{t}} e^{-\frac{|x-y|^2}{c_2 t}}.$$

Substituting $w = z - \beta\sqrt{t}$ we obtain

$$p_t^\beta(x, y) = \frac{1}{\sqrt{2\pi t}} \int_{\frac{|x-y|}{\sqrt{t}} - \beta\sqrt{t}}^{\infty} (w + \beta\sqrt{t}) e^{-\frac{w^2}{2}} dw \leq \frac{c_1}{\sqrt{t}} e^{-\frac{|x-y|^2}{c_2 t}}.$$

Set $|x - y| = \beta t$. Then

$$\int_0^\infty (w + \beta\sqrt{t}) e^{-\frac{w^2}{2}} dw \leq \sqrt{2\pi} c_1 e^{-\beta^2 t / c_2}.$$

This is not possible, since when $t \to \infty$ the left hand side grows to infinity, while the right hand side decreases to 0.

We close this section by proving that under Assumption 2.4 the transition kernel p_t satisfies uniform Gaussian upper bounds. To that purpose, we adapt the strategy of [75, Lemma 2]. We approximate $(\sigma, b) \in \Theta$ by a sequence of smooth coefficients, claim the uniform bounds for the approximating sequence and finally argue that the bound transfers to the limit.

Theorem 2.12. Consider σ, b satisfying Assumption 2.4 with σ weakly differentiable. Let X be the solution of the SDE (2.1). The transition kernel p_t of X satisfies

$$p_t(x, y) \leq \frac{C_T}{\sqrt{t}} e^{-\frac{(x-y)^2}{ct}}, \text{ for all } x, y \in [0, 1],\ 0 < t \leq T,$$

where the positive constants c, C_T depend only on λ, Λ.

Proof. Consider smooth σ_n, b_n satisfying Assumption 2.4 and such that

$$\lim_{n\to\infty} \|\sigma_n - \sigma\|_\infty = 0, \quad \lim_{n\to\infty} \|b_n - b\|_\infty = 0.$$

Let $X^{(n)}$ be a reflected diffusion process corresponding to (σ_n, b_n) and $p_t^{(n)}$ its transition density. The generator $L^{(n)}$ of $X^{(n)}$ is an elliptic divergence operator. Hence, the Aronson's and Nash estimates (see [85]) imply that for all $x, y \in [0,1]$, $t \leq T$

$$p_t^{(n)}(x,y) \leq \frac{C_T}{\sqrt{t}} e^{-\frac{(x-y)^2}{ct}},$$

where the constants C_T, c depend only on the uniform bounds on the coefficients b_n and σ_n. Since [44, Proposition 6.4] implies

$$\lim_{n\to\infty} \|p_t^{(n)}(\cdot,\cdot) - p_t(\cdot,\cdot)\|_\infty = 0,$$

we conclude that the claim holds. $\square$

3. Spectral estimation with random sampling times

This chapter is a modified version of the paper Chorowski and Trabs [23] which is accepted for publication in Stochastic Processes and their Applications. The nonparametric estimation of the volatility and the drift coefficient of a scalar diffusion is studied when the process is observed at random time points. The constructed estimator generalizes the spectral method by Gobet, Hoffmann and Reiß [44]. The estimation procedure is optimal in the minimax sense and adaptive with respect to the sampling time distribution and the regularity of the coefficients. The proofs are based on the eigenvalue problem for the generalized transition operator. The finite sample performance is illustrated in a numerical example.

3.1. Introduction

For decades diffusion models are used to describe the dynamics of continuous stochastic processes, for instance, stock prices in econometrics or particle movements in biology and physics. The statistical properties of diffusion models depend essentially on the observation scheme, where it is natural to assume discrete observations of the process. Mostly, equidistant observations are studied in the literature, distinguishing between high-frequent and low-frequent observations, depending whether the observation distance tends to zero or remains fixed. A summary of parametric methods is given by Aït-Sahalia [4]. Nonparametric estimation methods are surveyed by Fan [35].

As argued by Aït-Sahalia and Mykland [5], assuming equidistant observations might however not be realistic in many applications and random sampling times should be considered instead. For parametric estimation problems Aït-Sahalia and Mykland [5, 6] have shown that random sampling has a strong effect on the statistical problem and the performance of estimators. Naturally, the question arises how nonparametric estimators can be constructed for random sampling times and whether their (asymptotic) behavior is similar or worse than for equidistant observations.

In order to study the nonparametric estimation of the drift and the volatility coefficient of the diffusion when the process is observed at random times, we generalize the low-frequency results by Gobet et al. [44]. As they do, we consider a reflected scalar diffusion on a one-dimensional interval. By the compactness of the interval and the reflecting boundary, the diffusion is ergodic and admits a spectral gap. Our procedure relies on a representation of the coefficients in terms of the invariant measure and the first non-trivial eigenpair of the infinitesimal generator of the diffusion. This spectral identification method was introduced in Hansen et al. [46] and has been further studied by [20]. It is crucial that the eigenpair is determined by the transition operator of the time changed diffusion, where the time change is given by the Laplace transform of the sampling distribution. The former can be estimated by a wavelet projection method and latter by classical empirical process theory. As a side product of our analysis we clarify some aspects of the estimator and the proofs by Gobet et al. [44]. In particular, in order to stabilize the estimator against large stochastic errors a truncation with an in practice unknown threshold value was needed, which we could omit.

Moreover, we show that Lepski's method can be applied to choose the projection level in a data-driven way. This allows to adapt on the unknown Sobolev regularity of the drift and volatility coefficients of the diffusion. The first adaptive estimator based on low-frequency observations of a diffusion process has been constructed only recently in Söhl and Trabs [82]. Considering diffusion on the whole real line, this first result is restricted to a diffusion with constant volatility, simplifying the whole estimation problem, we do not need any additional restrictions on the drift or the volatility.

We prove that the estimators achieve minimax optimal convergence rates. The adaptive estimator only loses a logarithmic factor. In view of the *cost of randomness* determined by Aït-Sahalia and Mykland [6], it might be surprising that the convergence rates do not depend on the sampling distribution and coincide in fact with the nonparametric rates of the low-frequency setting. In that sense, our method is also adaptive with respect to the unknown sampling distribution. As one can see clearly from simulations, there is, however, a large *cost of ignoring the randomness* in the misspecified case where one applies the low-frequency estimator to randomly sampled observations using the average time step as observations distance.

The chapter is organized as follows: In Section 3.1.1 we introduce the diffusion with reflected boundaries, our basic assumptions and the main properties of the process. The estimators are constructed in Section 3.1.2. The main results on the convergence rates are stated and discussed in Section 3.2. The adaptive estimator is constructed in Section 3.3. The finite sample performance of the method is illustrated in a small simulation study in Section 3.4. The proofs of the upper and lower bounds as well as for the Lepski method are postponed to Sections 3.5, 3.6 and 3.7, respectively.

3.1.1. The model

Without loss of generality we can consider the unit interval $[0,1]$ for the reflecting diffusion. For a measurable and bounded drift function $b\colon [0,1]\to\mathbb{R}$ and a continuous volatility function $\sigma\colon [0,1]\to\mathbb{R}_+$ let the process $X=\{X_t : t\geq 0\}$ be given by the stochastic differential equation

$$\begin{aligned} dX_t &= b\left(X_t\right)dt+\sigma\left(X_t\right)dW_t+v\left(X_t\right)dK_t\left(X\right), \\ X_0 &= x_0, \text{ and for all } t\geq 0\ X_t\in[0,1], \end{aligned} \tag{3.1}$$

where x_0 is a random variable on $[0,1]$, $W=\{W_t : t\geq 0\}$ is a standard Brownian motion, $v\colon [0,1]\to\mathbb{R}$ satisfies $v(0)=1, v(1)=-1$, and K, which is part of the solution, is a non-anticipative continuous non-decreasing process increasing only when $X_t\in\{0,1\}$. By the Engelbert-Schmidt theorem boundedness of the drift coefficient together with the volatility function being continuous and strictly positive ensure that (3.1) has a weak solution, see Rozkosz and Słomiński [76, Thm. 4.1]. We denote by $\mathbb{P}_{\sigma,b}$ the law of this solution on the canonical space $\Omega=C(\mathbb{R}_+,[0,1])$ of continuous functions equipped with the topology of uniform convergence on compact subsets and endowed with its Borel $\sigma-$field $\mathcal{F}$.

For $N\in\mathbb{N}$ our observations are given by

$$(0,X_0),(\tau_1,X_{\tau_1}),\ldots,(\tau_N,X_{\tau_N})\in[0,\infty)\times[0,1]$$

where $\tau_1, \ldots, \tau_N$ is an increasing sequence of random time points. For convenience we write $\tau_0 = 0$.

Assumption 3.1. *Let the observation distances*

$$\Delta_n := \tau_n - \tau_{n-1}, \quad n = 1, \ldots, N,$$

be an independent and identically distributed sequence of strictly positive random variables with law

$$\gamma \in \Gamma := \Gamma(I, \alpha) := \{\gamma \text{ probability measure on } \mathbb{R}_+ : \gamma(I) \geq \alpha\}$$

for some compact interval $I \subset (0, \infty)$ and some $\alpha \in (0, 1]$. Let Δ_n be independent of the diffusion process X.

This condition on the sampling distributions is very weak. For every given positive distribution γ there are I, α such that $\gamma \in \Gamma(I, \alpha)$. The only restrictions are that the set Γ has to be bounded in the right sense, since we will derive uniform rates in this class, and we have to exclude distributions that concentrate at zero. The latter condition is natural because otherwise the observations would be of high-frequency type which would require a completely different analysis.

Example 3.2.

(i) The special case of the low-frequency observations is covered by setting $\tau_n = n\Delta$ for some fixed deterministic $\Delta > 0$. Then the sampling distribution is given by the Dirac measure in Δ, that is $\Gamma = \{\delta_\Delta\}$.

(ii) If the observation times are governed by a Poisson process, the waiting time to the next observation is exponentially distributed, that is $\gamma = Exp(\lambda)$ for some intensity $\lambda > 0$. In this case we can choose $\Gamma = \{Exp(\lambda) : \lambda \in \Lambda\}$ for any bounded set $\Lambda \subset (0, \infty)$.

To state the assumptions on the diffusion coefficients, we denote the $L^2([0, 1])$ *Sobolev space* of order $s > 0$ by $H^s := H^s([0, 1])$. Furthermore, let $H^s_b \subset H^s$ be the subset of bounded functions with Sobolev regularity s. Note that $H^s_b = H^s$ for $s > 1/2$ by the Sobolev embeddings.

Assumption 3.3. *For $s > 1$ and constants $d, D > 0$ let $(\sigma, b) \in \Theta_s$ where*

$$\Theta_s := \Theta_s(d, D) = \Big\{(\sigma, b) \in H^s \times H^{s-1}_b : \|\sigma^2\|_{H^s} \leq D, \ \|b\|_{H^{s-1}} \leq D, \inf_x \sigma(x) \geq d\Big\}.$$

In particular, $(\sigma, b) \in \Theta_s$ ensures the existence of a weak solution of (3.1). As shown by Gobet et al. [44] the compactness of $[0, 1]$ and the reflecting boundary conditions imply that X has a spectral gap and thus it is geometrically ergodic and admits an invariant measure μ. Focusing on asymptotic results, we can suppose that the initial value x_0 is distributed according to μ. Assumption 3.3 implies that μ has the Lebesgue density, abusing notation denoted by μ as well,

$$\mu(x) := \mu_{\sigma,b}(x) = C_0 \sigma^{-2}(x) \exp\Big(\int_0^x 2b(x)\sigma^{-2}(y)\, dy\Big), \quad x \in [0, 1], \tag{3.2}$$

for some normalizing constant $C_0 > 0$, cf. Bass [12, Chap. 4] or Karlin and Taylor [55, Chap. 15, Sect. 6]. It is easy to see that the regularity assumptions on b and σ imply that

$\mu \in H^s$, which will be essential for the analysis of the estimators. From the explicit formula for μ moreover follows that there are constants $0 < c < C$ such that $c \leq \mu_{\sigma,b} \leq C$ for any $(\sigma, b) \in \Theta_s$. Consequently, $L^2(\mu)$ with the inner product

$$\langle f, g \rangle_\mu := \int_0^1 f(x)g(x)\mu(x)dx$$

is a Hilbert space equivalent to $L^2([0,1])$.

Noting that reflection corresponds to Neumann boundary conditions, the infinitesimal generator $L = L_{\sigma,b}$ of the diffusion X is an unbounded, densely defined operator on $L^2([0,1])$ satisfying

$$\begin{aligned} Lf(x) &= b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \\ \mathrm{dom}(L) &= \left\{ f \in H^2([0,1]) : f'(0) = f'(1) = 0 \right\}. \end{aligned}$$

Furthermore, seen as an operator on the Hilbert space $L^2(\mu)$, the generator L is an elliptic, self-adjoint operator with compact resolvent, see Appendix A.2. Consequently, it has a pure point spectrum $\sigma(L) = \{v_k : k = 0, 1, ...\}$ and the corresponding eigenfunctions u_k form an $L^2(\mu)$ orthogonal basis. Its largest eigenvalue v_0 equals 0 with constant corresponding eigenfunction. All other eigenvalues are negative and we assume that they are ordered with respect to their multiplicities $0 > v_1 \geq v_2 \geq ...$. As shown in Proposition A.5, the eigenvalue v_1 is simple and the eigenfunction u_1 can be chosen strictly increasing.

3.1.2. Estimation method

The main idea used for the construction of the spectral estimators in [44] is that the coefficients of a stationary diffusion process can be expressed in terms of the invariant density μ and any nontrivial eigenpair (v_k, u_k), $k \geq 1$. Indeed, expressing the invariant measure in terms of the speed measure together with the Neumann boundary conditions yields, cf. [44, Sect. 3.1],

$$\sigma^2(x) = \frac{2v_k \int_0^x u_k(y)\mu(y)dy}{u_k'(x)\mu(x)}, \tag{3.3}$$

$$\begin{aligned} b(x) &= \frac{v_k u_k(x)}{u_k'(x)} - \frac{\sigma^2(x)u_k''(x)}{2u_k'(x)} \\ &= v_k \frac{u_k(x)u_k'(x)\mu(x) - u_k''(x)\int_0^x u_k(y)\mu(y)dy}{u_k'(x)^2\mu(x)}. \end{aligned} \tag{3.4}$$

Applying the ergodicity, it is easy to estimate the invariant measure μ. To recover an eigenpair of the generator, Gobet et al. [44] have used discrete equidistant observations, i.e. $\Delta_n = \Delta$ for some fixed $\Delta > 0$, to construct a matrix estimator of the transition operator $P_\Delta = e^{\Delta L}$. Noting that P_Δ shares eigenfunctions with the generator L while its eigenvalues are $e^{\Delta v_k}$, $k = 0, 1, ...$, they have obtained estimators of (v_k, u_k). We will generalize these results taking into account the random observation times $\tau_1, \ldots, \tau_N$.

Similar to the transition operator P_Δ we introduce the *generalized transition operator* R

on $L^2(\mu)$ given by

$$Rf(x) = \mathbb{E}_{\sigma,b,\gamma}\left[f(X_\tau)|X_0 = x\right], \quad x \in [0,1], \tag{3.5}$$

where τ is a random variable with distribution γ being independent of the process X. The crucial insight is that for any eigenpair (v_k, u_k) of the generator we have

$$Ru_k(x) = \mathbb{E}_\gamma[P_\tau u_k(x)] = \mathbb{E}_\gamma\left[e^{\tau v_k}\right] u_k(x) = \underbrace{\mathcal{L}_\gamma(-v_k)}_{=:\kappa_k} \cdot u_k(x), \tag{3.6}$$

where

$$\mathcal{L}_\gamma(z) := \int_0^\infty e^{-tz}\gamma(dt), \quad z \in \mathbb{R}_+, \tag{3.7}$$

is the Laplace transform of γ. Consequently, R is a compact operator with eigenvalues $1 = \kappa_0 > \kappa_1 > \kappa_2 \geq \kappa_3 \geq \ldots > 0$. In the functional calculus sense we obtain

$$R = \mathcal{L}_\gamma\left(-L\right).$$

Therefore, we can estimate the eigenpairs (v_k, u_k) using the spectral properties of R. Since the sampling distribution γ is unknown, we need to estimate the Laplace transform from the observations $(\Delta_n)_{n=1,\ldots,N}$.

Example 3.4 (Example 3.2 (continued))**.** (i) For $\Delta_n \equiv \Delta$ for some fixed $\Delta > 0$ we have $Rf = P_\Delta f$ and $\mathcal{L}_\gamma(z) = e^{-\Delta z}, z \geq 0$. We thus exactly recover the situation studied in [44].

(ii) If $\Delta_n \sim Exp(\lambda)$, then the Laplace transform is given by $\mathcal{L}_\gamma(z) = \int_0^\infty \lambda e^{-t(z+\lambda)}dt = \frac{\lambda}{z+\lambda}, z \geq 0$ and the operator R is the resolvent of the generator L.

The distribution of the eigenvalues of the operator R is inherited from the generator L and the sampling distribution γ. More precisely, we obtain the following lemma whose proof is postponed to Section 3.5.1.

Lemma 3.5. *Grant Assumptions 3.1 and 3.3. The spectral gap, that is* $\inf_{i\neq 1}|\kappa_i - \kappa_1|$, *and the eigenvalues of the generalized transition operator R have a lower bound uniform in* $(\sigma, b) \in \Theta_s$ *and* $\gamma \in \Gamma$.

3.1.3. Construction of the estimators

Let us fix some notation. We will write $f \lesssim g$ (resp. $g \gtrsim f$) when $f \leq C \cdot g$ for some universal constant $C > 0$. $f \sim g$ is equivalent to $f \lesssim g$ and $g \lesssim f$. Let (ψ_λ), with multi-indices $\lambda = (j,k)$, be an L^2-orthonormal regular wavelet basis of $L^2([0,1])$. The corresponding approximation spaces are given by

$$V_J := \overline{\operatorname{span}}\{\psi_\lambda : |\lambda| = |(j,k)| := j \leq J\}.$$

The L^2-orthogonal and the $L^2(\mu)-$orthogonal projections onto V_J are denoted by π_J and π_J^μ, respectively.

In fact, the approximation spaces do not necessarily need to be generated by wavelets. We only require that $V_J, J \in \mathbb{N}$, satisfy Jackson and Bernstein type inequalities with respect to the Sobolev spaces H^s, that is for all $0 \le t \le s$, $f \in H^s$ and $g \in V_J$

$$\|(I-\pi_J)f\|_{H^t} \lesssim 2^{-J(s-t)}\|f\|_{H^s} \quad \text{and} \quad \|g\|_{H^j} \lesssim 2^{Jj}\|g\|_{L^2}, \quad j=1,2, \tag{3.8}$$

and additionally we need the uniform bound

$$\Big\| \sum_{|\lambda|\le J} \psi_\lambda^2 \Big\|_\infty \lesssim \dim(V_J) = 2^J. \tag{3.9}$$

It follows from the well known properties of wavelets that (3.8) and (3.9) are satisfied.

Remark 3.6. Since the eigenfunctions of the generator of the reflected Brownian motion are given by the trigonometric functions, it seams particularly attractive to choose V_J as the trigonometric basis spanned by cosine functions.

After having fixed the basis functions and the corresponding approximation spaces V_J, there is a one-to-one correspondence between a linear operator $A\colon V_J \to V_J$ on the finite dimensional space V_J and its matrix representation $(A_{\lambda,\lambda'}) \in \mathbb{R}^{\dim V_J \times \dim V_J}$ with $A_{\lambda,\lambda'} := \langle \psi_\lambda, A\psi_{\lambda'}\rangle$. To simplify the notation, we will throughout use A to denote the operator as well as its representation matrix.

Using the ergodicity of the diffusion X and the independence of X and $(\Delta_n)_n$, the sequence $(X_{\tau_n})_n$ is ergodic, too. The natural estimator for the invariant measure is therefore the empirical measure

$$\mu_N = \frac{1}{N+1}\sum_{n=0}^{N} \delta_{X_{\tau_n}}.$$

To regularize μ_N, we define the projection estimator

$$\widehat{\mu}_J(x) := \sum_{|\lambda|\le J} \langle \psi_\lambda, \mu_N\rangle \psi_\lambda(x) \quad \text{with} \quad \langle \psi_\lambda, \mu_N\rangle := \frac{1}{N+1}\sum_{n=0}^{N}\psi_\lambda(X_{\tau_n})$$

for a projection level $J \in \mathbb{N}$. We proceed similarly to Gobet et al. [44]. Extending the matrix estimator of the transition semigroup, we introduce the matrix estimator $\widehat{R}_J = (\widehat{R}_{\lambda,\lambda'})$ of the action of the operator R from (3.5) on the wavelet basis with respect to the scalar product $\langle\cdot,\cdot\rangle_\mu$:

$$\widehat{R}_{\lambda,\lambda'} := \frac{1}{2N}\sum_{n=0}^{N-1}\Big(\psi_\lambda\left(X_{\tau_{n+1}}\right)\psi_{\lambda'}\left(X_{\tau_n}\right) + \psi_{\lambda'}\left(X_{\tau_{n+1}}\right)\psi_\lambda\left(X_{\tau_n}\right)\Big).$$

Since the observation times are independent of the diffusion, conditioning on τ_n, we can verify that $\widehat{R}_J$ is an unbiased estimator of the action of the operator R on the basis, that is

$$\mathbb{E}_{\sigma,b,\gamma}[\widehat{R}_{\lambda,\lambda'}] = \langle \psi_\lambda, R\psi_{\lambda'}\rangle_\mu.$$

The Gram matrix $G_J = (\langle \psi_\lambda, \psi_{\lambda'}\rangle_\mu)_{\lambda,\lambda'} \in \mathbb{R}^{\dim V_J \times \dim V_J}$ is determined by $\langle v, G_J v\rangle = \langle v, v\rangle_\mu$ for all $v \in V_J \setminus \{0\}$. Hence, G_J is a restriction of the scalar product $\langle\cdot,\cdot\rangle_\mu$ to finite dimensional space V_J. It can be estimated by $\widehat{G}_J = (\widehat{G}_{\lambda,\lambda'})$ with

$$\widehat{G}_{\lambda,\lambda'} = \frac{1}{N}\Big(\frac{1}{2}\psi_\lambda(X_0)\psi_{\lambda'}(X_0) + \sum_{n=1}^{N-1}\psi_\lambda(X_{\tau_n})\psi_{\lambda'}(X_{\tau_n}) + \frac{1}{2}\psi_\lambda(X_{\tau_N})\psi_{\lambda'}(X_{\tau_N})\Big),$$

satisfying

$$\mathbb{E}_{\sigma,b,\gamma}[\widehat{G}_{\lambda,\lambda'}] = \langle \psi_\lambda, \psi_{\lambda'} \rangle_\mu = \langle \psi_\lambda, G_J \psi_{\lambda'} \rangle.$$

Owing to $\langle v, G_J v\rangle = \langle v, v\rangle_\mu > 0$ for any $v \in V_J \setminus \{0\}$, the matrix G_J is invertible. By construction $\langle v, \widehat{G}_J v\rangle$ is always non-negative and it will be strictly positive whenever the sample is sufficiently dispersed over all the interval $[0, 1]$. By ergodicity we can expect this to be a high probability event. With a Neumann series argument we can moreover bound the norm of $\widehat{G}_J^{-1}$ as stated by the following lemma, which is proven in Section 3.5.4.

Lemma 3.7. *Grant Assumption 3.1 and 3.3. On the event* $\mathcal{T}_1 = \left\{ \|G_J - \widehat{G}_J\|_{L^2} \le \frac{1}{2}\|G_J^{-1}\|_{L^2}^{-1} \right\}$ *the estimator* $\widehat{G}_J$ *is invertible and satisfies* $\|\widehat{G}_J{}^{-1}\|_{L^2} \le 2\|G_J^{-1}\|_{L^2}$. *Moreover,* $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_1) \le N^{-1}2^{2J}$ *holds uniformly over* Θ_s *and* Γ.

Whenever $\widehat{G}_J^{-1}$ exists, we can consider $\widehat{G}_J^{-1}\widehat{R}_J$. Since $\widehat{R}_J$ is symmetric it immediately follows that $\widehat{G}_J^{-1}\widehat{R}_J$ is symmetric with respect to the $\widehat{G}_J$-scalar product. Furthermore, by the Cauchy-Schwarz inequality and the inequality between geometric and arithmetic means we obtain for all $v \in V_J \setminus \{0\}$

$$\begin{aligned} \langle \widehat{R}_J v, v\rangle &= \frac{1}{N}\sum_{n=0}^{N-1} v\left(X_{\tau_n}\right) v\left(X_{\tau_{n+1}}\right) \\ &\le \frac{1}{N}\Big(\sum_{n=0}^{N-1} v^2(X_{\tau_n})\Big)^{1/2}\Big(\sum_{n=1}^{N} v^2(X_{\tau_n})\Big)^{1/2} \\ &\le \frac{1}{N}\Big(\frac{1}{2}v^2\left(X_0\right) + \frac{1}{2}v^2\left(X_{\tau_N}\right) + \sum_{n=1}^{N-1} v^2\left(X_{\tau_n}\right)\Big) = \langle \widehat{G}_J v, v\rangle. \end{aligned}$$

Consequently, all eigenvalues of the matrix $\widehat{G}_J^{-1}\widehat{R}_J$ are real and smaller than one. It is easy to check that 1 is an eigenvalue corresponding to the constant function. We define the estimator $(\widehat{\kappa}_{J,1}, \widehat{u}_{J,1})$ of the eigenpair (κ_1, u_1) as the eigenpair of the matrix $\widehat{G}_J^{-1}\widehat{R}_J$ corresponding to the biggest eigenvalue smaller than one. On the exceptional event that $\widehat{G}_J$ is not invertible, we set $\widehat{\kappa}_{J,1} = 0$ and $\widehat{u}_{J,1} = 1$. Furthermore we choose the estimated eigenfunction $\widehat{u}_{J,1}$ normalized in L^2.

Using $\widehat{\kappa}_{J,1}$ and the identification equation $\kappa_1 = \mathcal{L}_\gamma(-v_1)$, we can estimate v_1. The canonical estimator for the Laplace transform of γ is the Laplace transform of the empirical measure of the sampling distances $\Delta_n = \tau_n - \tau_{n-1}, n = 1, \ldots, N$. Hence, we define

$$\widehat{\mathcal{L}}(y) := \frac{1}{N}\sum_{n=1}^{N} e^{-y\Delta_n}, \quad y \in \mathbb{R}_+.$$

Due to the i.i.d. structure of (Δ_n), the classical empirical process theory shows that $\widehat{\mathcal{L}}$ estimates $\mathcal{L}_\gamma$ uniformly in a neighborhood of v_1 with the parametric rate $N^{-1/2}$. Moreover, $\widehat{\mathcal{L}}$ is strictly decreasing and continuous, thus invertible. We define

$$\widehat{v}_{J,1} := -\widehat{\mathcal{L}}^{-1}\left(\widehat{\kappa}_{J,1}\right) \mathbf{1}_{\{\widehat{\kappa}_{J,1} > 0\}}. \tag{3.10}$$

With the above definitions and in view of the identification formulas (3.3) and (3.4) we can define the plug-in estimators of the diffusion coefficients. In order to ensure integrability of our estimators, we need to stabilize against large stochastic errors. Using the prior knowledge that $(\sigma, b) \in \Theta_s$, especially $\|\sigma^2\|_\infty \leq D$ and $\|b\|_{L^2} \leq D$ for some $D > 0$, we thus define

$$\widehat{\sigma}_J^2(x) = 2\widehat{v}_{J,1} \frac{\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widehat{u}'_{J,1}(x)\widehat{\mu}_J(x)} \wedge D, \tag{3.11}$$

$$\widehat{b}_J(x) = \widetilde{b}_J(x)\mathbf{1}_{\{\|\widetilde{b}_J\|_{L^2} \leq 2D\}} \quad \text{for} \quad \widetilde{b}_J(x) := \frac{\widehat{v}_{J,1}\widehat{u}_{J,1}(x)}{\widehat{u}'_{J,1}(x)} - \frac{\widehat{\sigma}_J^2(x)\widehat{u}''_{J,1}(x)}{2\widehat{u}'_{J,1}(x)}. \tag{3.12}$$

3.2. Minimax convergence rates

Let us now state our first main results, generalizing Theorems 2.4 and 2.5 in [44], respectively. Note that since $u_1'(0) = u_1'(1) = 0$, the function

$$[0,1] \ni x \mapsto \frac{2v_1 \int_0^x u_1(y)\mu(y)dy}{u_1'(x)\mu(x)} = \sigma^2(x)$$

is defined in $\{0,1\}$ via continuous extension such that the proposed estimators $\widehat{\sigma}_J^2$ and $\widehat{b}_J$ might be unstable at the boundary. We restrict the L^2-loss to an interval $[a,b] \subset [0,1]$ for $0 < a < b < 1$ and refer to [44, Section 3.3.8] for a discussion of the boundary problem.

Theorem 3.8. Grant Assumptions 3.1 and 3.3 for some $s > 1$. Let $0 < a < b < 1$. Choosing $2^J \sim N^{1/(2s+3)}$, we have

$$\begin{aligned} \sup_{(\sigma,b,\gamma)\in\Theta_s\times\Gamma} \mathbb{E}_{\sigma,b,\gamma}\big[\|\widehat{\sigma}_J^2 - \sigma^2\|^2_{L^2([a,b])}\big] &\lesssim N^{-2s/(2s+3)}, \\ \sup_{(\sigma,b,\gamma)\in\Theta_s\times\Gamma} \mathbb{E}_{\sigma,b,\gamma}\big[\|\widehat{b}_J - b\|^2_{L^2([a,b])}\big] &\lesssim N^{-2(s-1)/(2s+3)}. \end{aligned}$$

The risk of $\widehat{\sigma}^2$ and $\widehat{b}$ decomposes into the errors for estimating the invariant density μ and the eigenpair and (v_1, u_1) of the infinitesimal generator L of the diffusion. In view of formula (3.2) the invariant density inherits Sobolev regularity of degree s from the diffusion coefficients. Together with the ergodicity and the spectral gap μ can be estimated with the rate $\mathbb{E}_{\sigma,b,\gamma}[\|\widehat{\mu}_J - \mu\|_{L^2}] \lesssim N^{-\frac{s}{2s+1}}$ if we choose $2^J \sim N^{-1/(2s+1)}$, cf. Proposition 3.12. Due to $\mathcal{L}_\gamma(-v_1) = \kappa_1$ estimating v_1 reduces to estimate the eigenvalue κ_1 of the operator R and the inverse of the Laplace transform $\mathcal{L}_\gamma$ in a neighborhood of κ_1. The latter estimation problem can be solved with standard empirical process results yielding the parametric rate $N^{-1/2}$ for $\widehat{\mathcal{L}}$, see Lemma 3.19.

The analysis of the estimation error of the eigenpair (κ_1, u_1) of the generalized transition operator R is the most challenging ingredient of our proofs. We first restrict the eigenvalue problem to the finite dimensional space V_J, that is we find $(\kappa_{J,1}, u_{J,1}) \in \mathbb{R}_+ \times V_J$ such that

$$\langle v, Ru_{J,1}\rangle_\mu = \kappa_{J,1}\langle v, u_{J,1}\rangle_\mu \quad \text{for all } v \in V_J. \tag{3.13}$$

As shown in Theorem A.1 the resulting approximation error $\|u_1 - u_{J,1}\|_{L^2(\mu)} + |\kappa_1 - \kappa_{J,1}|$ is controlled by the spectral gap of the operator R and the smoothness of the eigenfunction (of degree $s+1$) achieving the order of magnitude $2^{-J(s+1)}$. In the second step we approximate the finite dimensional problem (3.13) by a generalized symmetric eigenvalue problem for the random matrices $\widehat{R}_J$ and $\widehat{G}_J$. We use classical a posteriori error bounds to show that the approximation error is controlled by the norm of the so called residual vector $r = (\widehat{R}_J - \kappa_{J,1}\widehat{G}_J)u_{J,1}$, cf. Theorem A.10. $\|r\|_{L^2}$ can be bounded by the matrix approximation errors $\|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2}$ and $\|(\widehat{G}_J - G_J)u_{J,1}\|_{L^2}$ that tend to zero by the mixing property of the Markov chain $(X_{\tau_n})_n$. A delicate point is that the a posteriori technique gives an existence statement, but does not bound the error between ordered eigenpairs. We overcome this difficulty using the absolute Weyl theorem for generalized symmetric eigenvalue problems, see [62]. We conclude that (κ_1, u_1) can be estimated with the rate $N^{-(s+1)/(2s+3)}$.

Because the volatility estimator relies on the first derivative of the eigenfunction the statistical problem is ill-posed of degree one, deteriorating the rate to $N^{-s/(2s+3)}$. For the drift estimator we need the second derivative, adding a degree of ill-posedness. At the same time the regularity of b is smaller such that the rate becomes $N^{-(s-1)/(2s+3)} = N^{-(s-1)/(2(s-1)+5)}$. Compared to Gobet et al. [44], the same rates can thus be achieved with random sampling times (with unknown sampling distribution) than with equidistant low frequent observations. In fact, the convergence rates are optimal in the minimax sense:

Theorem 3.9. Grant Assumption 3.1 for an arbitrary $\gamma \in \Gamma$ admitting a bounded Lebesgue density at the origin. Grant Assumption 3.3 for some $s > 1$. For $0 < a < b < 1$ it holds

$$\begin{aligned}
\inf_{\bar{\sigma}} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b,\gamma}\big[\|\bar{\sigma}^2 - \sigma^2\|^2_{L^2([a,b])}\big] &\gtrsim N^{-2s/(2s+3)}, \\
\inf_{\bar{b}} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b,\gamma}\big[\|\bar{b} - b\|^2_{L^2([a,b])}\big] &\gtrsim N^{-2(s-1)/(2s+3)},
\end{aligned}$$

where the infimum is taken over all estimators, i.e. measurable functions, $\bar{\sigma}$ and $\bar{b}$, respectively.

The proof of the lower bounds for observations sampled at random times follows the same strategy as for low frequency observations in [44]. Constructing alternatives that admit the same invariant measure, proving the lower bound is reduced to a testing problem by Assouad's lemma, see Tsybakov [89, Sect. 2.7.2]. The Kullback-Leibler distance between the distributions of two alternatives can then be bounded in terms of the L^2-distance between the kernels of the corresponding operators R from (3.5), which is finally accomplished using Hilbert-Schmidt norm estimates and the explicit form of the inverse of the generator.

3.3. Adaptive estimation

The optimal choice of the projection level crucially depends on the unknown smoothness s. In this section, we construct a completely data driven estimation procedure adapting to the Sobolev regularity of σ^2 and b. We focus on the volatility estimator, noting that the methodology should extend to the drift estimation without additional theoretical problems. We adopt the general adaptation principle by Lepski [60].

The aim is to choose the optimal projection level from the set

$$\mathcal{J}_N := [J_{min}, J_{max}] \cap \mathbb{N} \quad \text{with} \quad 2^{J_{min}} \sim \log N, \quad 2^{J_{max}} \sim \tfrac{N}{(\log N)^2 \log\log N}.$$

For any $J \in \mathcal{J}_N$ we define

$$s_J^2 := \Lambda^2 2^{3J} \frac{\log\log N}{N}, \tag{3.14}$$

for some appropriate constant $\Lambda > 0$ depending on d, D as well as I, α (but not on s) from Assumptions 3.1 and 3.3. The quantity s_J is an upper bound for the stochastic error of $\widehat{\sigma}^2_J$, cf. Corollary 3.25. The adaptive estimator is defined by

$$\widetilde{\sigma}^2 := \widehat{\sigma}^2_{\widehat{J}} \quad \text{with} \quad \widehat{J} := \min\left\{J \in \mathcal{J}_N : \forall_{K \geq J, K \in \mathcal{J}_N} \|\widehat{\sigma}^2_K - \widehat{\sigma}^2_J\|_{L^2([a,b])} \leq s_K\right\}.$$

Heuristically, $\widehat{J}$ is the smallest projection level for which the stochastic error still dominates the bias.

Our main result for the adaptive estimation shows that the estimator $\widetilde{\sigma}^2$ achieves the optimal convergence rate up to an additional $\log \log N$ factor.

Theorem 3.10. Grant Assumption 3.1 and define $\Gamma_0 := \{\gamma \in \Gamma : \mathbb{E}_\gamma[\tau^{-1/2}] \leq D\}$. Let Assumption 3.3 be fulfilled for some $s > 5/2$. Let $0 < a < b < 1$. Then, there exists for every $\varepsilon > 0$ some $C > 0$ such that, for N sufficiently large, we have

$$\sup_{(\sigma, b, \gamma) \in \Theta_s \times \Gamma_0} \mathbb{P}_{\sigma,b,\gamma}\Big(\|\widetilde{\sigma}^2 - \sigma^2\|^2_{L^2([a,b])} > C\Big(\frac{\log \log N}{N}\Big)^{2s/(2s+3)}\Big) < \varepsilon.$$

The proof of this theorem is postponed to Section 3.7. It relies on a concentration inequality for the Markov chain $(X_{\tau_n})_{n\geq 0}$, see Proposition 3.24 as well as Nickl and Söhl [65, Section 3]. For the latter we need the additional assumption on γ allowing for a uniform bound on the transition density of the time-changed diffusion process. Up to the concentration result, the proof relies on the standard arguments for the Lepski method.

3.4. Numerical example

In this section, we present numerical results for the volatility estimation. Throughout the chapter, we consider a diffusion process X with linear mean reverting drift $b(x) = 0.2 - 0.4x$, quadratic squared volatility function $\sigma^2(x) = 0.4 - (x - 0.5)^2$ and two reflecting barriers at 0 and 1. The sample paths were generated using Euler-Maruyama scheme with time step size 0.001 and reflection after each step. The source code used for performing the simulations is available at [22].

For $\Delta = 0.25$ we compare the estimation error for four different sampling distributions of quite different shapes: the case of equidistant observations with frequency Δ^{-1}, the uniform distribution on the interval $[0, 2\Delta]$, the symmetric Beta$(0.2, 0.2)$ distribution rescaled to the interval $[0, 2\Delta]$ and finally, the exponential distribution with intensity Δ^{-1}. Note that all considered distributions have mean Δ, Uniform and Beta distributions have the same compact support $[0, 2\Delta]$ and together with exponential distribution they allow for arbitrary small sampling distances. Figure 3.1 depicts a fragment of a simulated trajectory of the diffusion together with the observations from different sampling schemes.

To construct the approximation spaces, we used the Fourier orthogonal cosines basis i.e.

$$V_J = \overline{\operatorname{span}}\{\sqrt{2}\cos(j\pi x) : 0 \leq j \leq J\},$$

see (3.6). We compare an oracle choice of the projection level with the adaptive estimator. As target interval we choose $[0.1, 0.9]$.

In Table 3.1 we compare the oracle and adaptive root mean integrated squared error

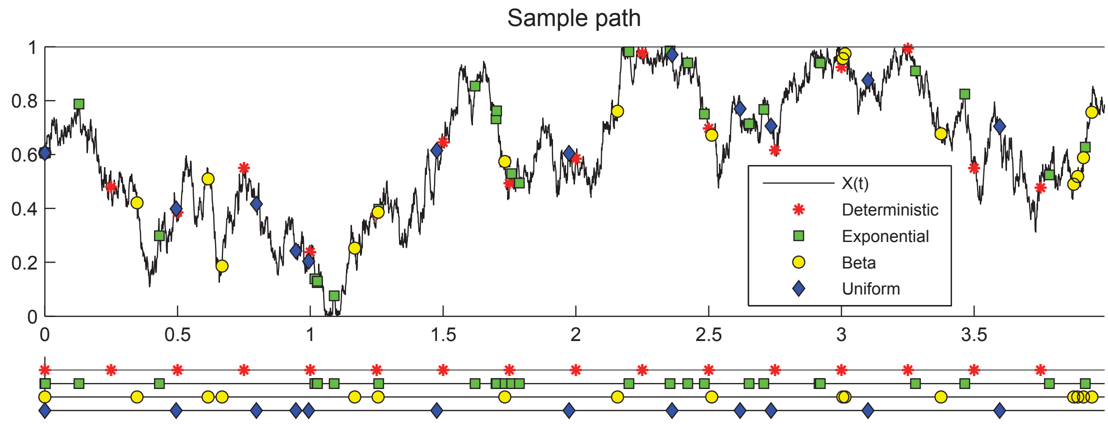

Figure 3.1.: Sample path of the process X for $0 \leq t \leq 4$ with marked observations from different sampling distributions.

Distribution \ Sample Size	Oracle projection level			Adaptive estimator		
	4 000	12 000	20 000	4 000	12 000	20 000
Deterministic	0.0233	0.0155	0.0123	0.0318	0.0214	0.0130
Uniform	0.0258	0.0168	0.0134	0.0341	0.0221	0.0139
Exponential	0.0282	0.0177	0.0141	0.0362	0.0231	0.0148
Beta	0.0296	0.0211	0.0179	0.0432	0.0255	0.0178

Table 3.1.: Root mean integrated squared error for volatility estimation on $[0.1, 0.9]$ based on 1000 Monte Carlo iterations.

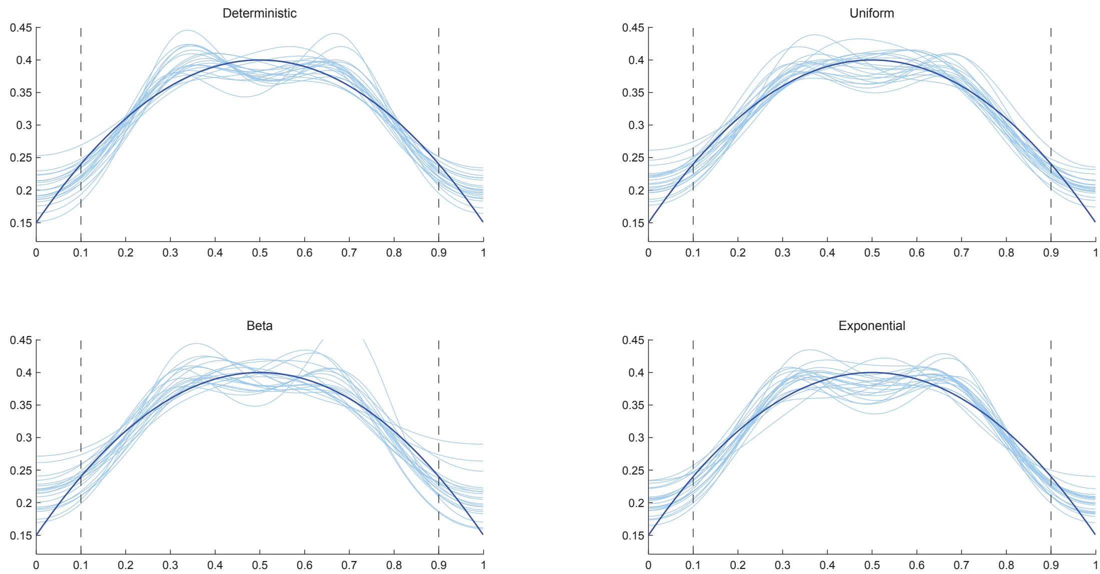

Figure 3.2.: Estimated volatility functions using the adapted estimator for 20 independent trajectories of the diffusion and four different sampling distributions with sample size N = 20 000.

(RMISE) for volatility estimation on the interval $[0.1, 0.9]$, obtained by a Monte Carlo simulation with 1000 iterations. The oracle projection level J is stable with respect to the sampling distribution and surprisingly small, taking values 2 for N = 4 000 and 4 for N = 12 000 and N = 20 000 across all distributions, with the exception of Beta with sample size N = 12 000, when it equals 2. For the adaptive estimation we chose the constant Λ in (3.14) equal to 0.01.

Relative to $\|\sigma^2\|_{L^2([0.1,0.9])} \approx 0.31$ the error of the oracle decreases from approximately 10% for sample size N = 4 000 to 5% for N = 20 000. In particular, for large sample size the error of the adaptive procedure is fairly close to the oracle error. The errors are quite stable across sampling distributions as the estimator, where the deterministic sampling allows for the smallest error and the Beta distribution generates the largest errors. The latter is not surprising because the Beta distribution is chosen in a way that yields a strong clustering of the observations.

For 20 independent paths and sample size N = 20 000 the resulting adaptive volatility estimators are shown in Figure 3.2. While the estimators behave nicely in the interior of the interval, the boundary problem outside the interval $[0.1, 0.9]$ is clearly visible. Again we see that the estimation for the Beta sampling distribution is the worst.

In the misspecified case where the randomness of the observation times is ignored, the RMISE of the low-frequency estimator designed for equidistant observations with Δ set to the average observation distance is four times larger than the error of our method in our simulations.

3.5. Proofs of the upper bounds

Throughout we take Assumptions 3.1 and 3.3 for granted.

3.5.1. Spectral properties of the generalized transition operator R

Recall that u_1 is the eigenfunction corresponding to the biggest negative eigenvalue v_1 of the generator L, normalized in $L^2([0,1])$. By Proposition A.5 u_1 can be chosen to be increasing and for any $0 < a < b < 1$ there exists a positive constant $c_{a,b} > 0$ such that

$$\inf_{(\sigma,b)\in\Theta_s} \inf_{x\in[a,b]} u_1'(x) > c_{a,b}. \tag{3.15}$$

By Proposition A.6 the family of generators $\{L_{\sigma,b} : (\sigma,b) \in \Theta_s\}$ has a uniform spectral gap on Θ_s, meaning that there is a constant $s_0 > 0$ such that

$$\inf_{(\sigma,b)\in\Theta_s} \inf_{i\neq 1} |v_i - v_1| = \inf_{(\sigma,b)\in\Theta_s} \{|v_1|, |v_2 - v_1|\} \geq s_0. \tag{3.16}$$

Moreover, arguing as in [29, Example 4.6.1] the eigenvalues v_k satisfy uniformly on Θ_s

$$C_1 k^2 \leq -v_k \leq C_2 k^2, \tag{3.17}$$

for constants $0 < C_1 < C_2$, while corresponding eigenfunctions u_k belong to the Sobolev space H^{s+1} (see [44, Lemma 6.6], c.f. Proposition A.6) fulfilling

$$\|u_k\|_{H^{s+1}} \lesssim (1 \vee |v_k|)^{\lceil s\rceil}. \tag{3.18}$$

As announced in Lemma (3.5) these bounds transfer uniformly to the operator R.

Proof of Lemma 3.5. For convenience we define $m := \min I > 0$ and $M := \max I$. By the definition of R and the uniform bounds on the eigenvalues v_k of L in (3.17), we have

$$\kappa_k = \mathcal{L}_\gamma(-v_k) = \int_0^\infty e^{tv_k}\gamma(dt) \geq \int_0^\infty e^{-tC_2k^2}\gamma(dt) \geq \alpha e^{-MC_2k^2} \quad \text{for } k \geq 1.$$

The spectral gap of the operator R equals $\min\{1-\kappa_1, \kappa_1-\kappa_2\}$. Due to (3.16), we have

$$\begin{aligned} \kappa_1 - \kappa_2 &= \int_0^\infty (e^{tv_1} - e^{tv_2})\gamma(dt) = \int_0^\infty e^{tv_2}(e^{t(v_1-v_2)} - 1)\gamma(dt) \\ &\geq \int_0^\infty e^{-4tC_2}(e^{ts_0} - 1)\gamma(dt) \geq \alpha e^{-4MC_2}(e^{ms_0} - 1). \end{aligned}$$

Similarly $1 - \kappa_1 = \int_0^\infty (1 - e^{tv_1})\gamma(dt) \geq \int_0^\infty (1 - e^{-tC_1})\gamma(dt) \geq \alpha(1 - e^{-mC_1})$. $\square$

3.5.2. Consequences of the mixing property

First, we establish general bounds for the variance of integrals with respect to the empirical measure which are due to the mixing behavior of the sequence $(X_{\tau_k})_k$. The following lemma is a straightforward generalization of [44, Lemma 6.2]. Since this is the key result to bound the stochastic error, we give the proof to keep the paper self-contained.

Lemma 3.11. *For bounded $H_1, H_2 \in L^2([0,1])$ we have the following two variance estimates:*

$$\begin{aligned}\operatorname{Var}_{\sigma,b,\gamma}\left[\frac{1}{N}\sum_{n=1}^{N} H_1(X_{\tau_n})\right] &\lesssim N^{-1}\mathbb{E}_{\sigma,b,\gamma}\left[H_1^2(X_0)\right],\\ \operatorname{Var}_{\sigma,b,\gamma}\left[\frac{1}{N}\sum_{n=0}^{N-1} H_1(X_{\tau_n})H_2(X_{\tau_{n+1}})\right] &\lesssim N^{-1}\mathbb{E}_{\sigma,b,\gamma}[H_1^2(X_0)H_2^2(X_{\tau_1})].\end{aligned}$$

Proof. Denote $f(X_{\tau_n}) = H_1(X_{\tau_n}) - \mathbb{E}_{\sigma,b,\gamma}[H_1(X_{\tau_n})]$. Consider $m \geq n$ and let $k = m-n$. Since process X is stationary and has a uniform spectral gap $\|R^k f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)}\mathcal{L}_\gamma^k(s_0)$ holds for every function f that is $L^2(\mu)$- orthogonal to constants. Arguing analogously as in the proof of Lemma 3.5we obtain $\sup_{\gamma\in\Gamma}\mathcal{L}_\gamma(s_0) < 1$. Hence, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}_{\sigma,b,\gamma}[f(X_{\tau_n})f(X_{\tau_m})] &= \mathbb{E}_{\sigma,b,\gamma}\left[f(X_{\tau_n})\mathbb{E}_{\sigma,b,\gamma}\left[f(X_{\tau_{n+k}})\,|X_{\tau_n}\right]\right]\\ &= \langle f, R^k f\rangle_\mu \leq \|f\|_{L^2(\mu)}^2\mathcal{L}_\gamma^k(s_0).\end{aligned}$$

Since $\|f\|_{L^2(\mu)}^2 = \operatorname{Var}_{\sigma,b,\gamma}[H_1(X_0)] \leq \mathbb{E}_{\sigma,b,\gamma}\left[H_1^2(X_0)\right]$ and

$$\operatorname{Var}_{\sigma,b,\gamma}\left[\sum_{n=1}^{N} H_1(X_{\tau_n})\right] = \sum_{n,m=1}^{N}\mathbb{E}_{\sigma,b,\gamma}[f(X_{\tau_n})f(X_{\tau_m})] \leq \|f\|_{L^2(\mu)}^2\sum_{n,m=1}^{N}\mathcal{L}_\gamma^{|n-m|}(s_0)$$

to prove the first inequality we just have to show that $\sum_{n,m=1}^{N}\mathcal{L}_\gamma^{|n-m|}(s_0) \lesssim N$. This easily follows from the formula for the sum of finite geometric series.

To prove the second inequality, first note that

$$\begin{aligned}&\operatorname{Var}_{\sigma,b,\gamma}\left[\frac{1}{N}\sum_{n=0}^{N-1} H_1(X_{\tau_n})H_2(X_{\tau_{n+1}})\right]\\ &\quad\leq \frac{1}{N^2}\mathbb{E}_{\sigma,b,\gamma}\Big[\sum_{n,m=0}^{N-1} H_1(X_{\tau_n})H_2(X_{\tau_{n+1}})H_1(X_{\tau_m})H_2(X_{\tau_{m+1}})\Big]\\ &\quad= \frac{1}{N^2}\mathbb{E}_{\sigma,b,\gamma}\Big[\sum_{n,m=0}^{N-1} H_1(X_{\tau_n})H_2(X_{\tau_{n+1}})H_1(X_{\tau_m})H_2(X_{\tau_{m+1}})\Big] - \langle H_1, RH_2\rangle_\mu^2.\end{aligned}$$

Since the sum of diagonal terms equals $N^{-1}\mathbb{E}_{\sigma,b,\gamma}\Big[H_1^2(X_0)H_2^2(X_{\tau_1})\Big]$, it does not exceed the claimed upper bound. The sum of the other terms equals

$$\frac{1}{N^2}\sum_{\substack{n,m=0\\ n\neq m}}^{N-1}\langle H_2\cdot(RH_1), R^{|n-m|-1}(H_1\cdot(RH_2) - \langle H_1, RH_2\rangle_\mu)\rangle_\mu \qquad \underbrace{-\frac{1}{N}\langle H_1, RH_2\rangle_\mu^2}_{\lesssim N^{-1}\mathbb{E}_{\sigma,b,\gamma}\left[H_1^2(X_0)H_2^2(X_{\tau_1})\right]}.$$

Using the spectral gap of the operator R together with the Cauchy-Schwarz inequality, we obtain that

$$\Big\|R^{|n-m|-1}(H_1\cdot(RH_2) - \langle H_1, RH_2\rangle_\mu)\Big\|_{L^2(\mu)} \lesssim \|H_1\cdot(RH_2)\|_{L^2(\mu)}\mathcal{L}_\gamma^{|n-m|-1}(s_0).$$

Consequently, using again Cauchy-Schwarz and the formula for the sum of finite geometric series, we can bound the considered variance by

$$\begin{aligned}
&\frac{1}{N^2}\sum_{\substack{n,m=0\\ n\neq m}}^{N-1}\|H_2\cdot(RH_1)\|_{L^2(\mu)}\|H_1\cdot(RH_2)\|_{L^2(\mu)}\mathcal{L}_\gamma^{|n-m|-1}(s_0)\\
&\qquad\lesssim\frac{1}{N}\|H_2\cdot(RH_1)\|_{L^2(\mu)}\|H_1\cdot(RH_2)\|_{L^2(\mu)}\\
&\qquad\lesssim\frac{1}{N}\mathbb{E}_{\sigma,b,\gamma}\left[H_2^2(X_0)H_1^2(X_{\tau_1})\right]^{1/2}\mathbb{E}_{\sigma,b,\gamma}\left[H_1^2(X_0)H_2^2(X_{\tau_1})\right]^{1/2}\\
&\qquad=\frac{1}{N}\mathbb{E}_{\sigma,b,\gamma}\left[H_2^2(X_0)H_1^2(X_{\tau_1})\right].
\end{aligned}$$

□

The first consequence of the previous result is the following bound for the risk of the estimator of the invariant measure.

Proposition 3.12. *Under Assumption 3.3 it holds*

$$\mathbb{E}_{\sigma,b,\gamma}\Big[\,\|\mu-\widehat{\mu}_J\|_{L^2}^2\,\Big]\lesssim N^{-2Js}+N^{-1}2^J. \tag{3.19}$$

Furthermore if we choose $2^J\sim N^{1/(2s+3)}$ *the event* $\mathcal{T}_0=\{\forall x\in[0,1]\,\inf\mu/2\leq\widehat{\mu}_J(x)\leq 2\sup\mu\}$ *satisfies* $\mathbb{P}_{\sigma,b,\gamma}\Big(\Omega\setminus\mathcal{T}_0\Big)\lesssim N^{-\frac{2s}{2s+3}}$.

Proof. The explicit formula (3.2) for μ shows that $\|\mu\|_{H^s}$ is uniformly bounded over Θ_s. Jackson's inequality yields

$$\|(I-\pi_J)\mu\|_{L^2}^2\lesssim 2^{-2Js}.$$

Using Lemma 3.11, we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b,\gamma}\left[\|\pi_J\mu-\widehat{\mu}_J\|_{L^2}^2\right] &= \sum_{|\lambda|\leq J}\mathbb{E}_{\sigma,b,\gamma}[\langle\psi_\lambda,\mu-\mu_N\rangle^2]=\sum_{|\lambda|\leq J}\mathrm{Var}_{\sigma,b,\gamma}\left[\langle\psi_\lambda,\mu_N\rangle\right]\\
&\lesssim N^{-1}\sum_{|\lambda|\leq J}\mathbb{E}_{\sigma,b,\gamma}\left[\psi_\lambda^2(X_0)\right]\lesssim 2^JN^{-1}
\end{aligned}$$

and (3.19) follows by the triangle inequality. Furthermore, by Jackson's inequality,

$$\begin{aligned}
\sup_{x\in[0,1]}\pi_J\mu(x) &\leq \|\mu\|_\infty+\|(I-\pi_J)\mu\|_\infty\lesssim\|\mu\|_{H^1}+\|(I-\pi_J)\mu\|_{H^1}\lesssim 1+2^{-J(s-1)}\\
\inf_{x\in[0,1]}\pi_J\mu(x) &\geq \inf_{x\in[0,1]}\mu(x)-\|(I-\pi_J)\mu\|_\infty\gtrsim 1-2^{-J(s-1)}.
\end{aligned}$$

Hence, for J large enough, $\pi_J\mu$ is bounded by $\frac{3}{4}\inf\mu$ from below and $\frac{3}{2}\sup\mu$ from above. Consequently, $\widehat{\mu}_J(x)$ lies in $[\frac{1}{2}\inf\mu,2\sup\mu]$ if $\|\widehat{\mu}_J-\pi_J\mu\|_\infty$ is small enough. For a given constant $C>0$, Bernstein's inequality shows

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}\Big(\|\widehat{\mu}_J-\pi_J\mu\|_\infty>C\Big)&\leq C^{-2}\mathbb{E}_{\sigma,b,\gamma}[\|\pi_J\mu-\widehat{\mu}_J\|_\infty^2]\lesssim\mathbb{E}_{\sigma,b,\gamma}[\|\pi_J\mu-\widehat{\mu}_J\|_{H^1}^2]\\
&\lesssim 2^{2J}\mathbb{E}_{\sigma,b,\gamma}\Big[\|\pi_J\mu-\widehat{\mu}_J\|_{L^2}^2\Big]\lesssim N^{-\frac{2s}{2s+3}}.
\end{aligned}$$

□

3.5.3. Analysis of the projection error

Denote by $(\kappa_{J,i}, u_{J,i})$, $i = 0, 1, 2, ..., \dim V_J - 1$, the eigenpairs of the operator $\pi_J^\mu R \pi_J^\mu$ ordered decreasingly with respect to the eigenvalues. Note that $(\kappa_{J,i}, u_{J,i})$ are solutions of the eigenvalue problem for the operator R restricted to the finite approximation spaces V_J on $L^2(\mu)$:

$$\langle R u_{J,i}, v\rangle_\mu = \kappa_{J,i}\langle u_{J,i}, v\rangle_\mu, \text{ for every } v \in V_J. \tag{3.20}$$

Take $u_{J,i}$ normalized in the L^2 norm. Since $\pi_J^\mu R \pi_J^\mu$ is a positive definite self-adjoint operator on $L^2(\mu)$ with $\|\pi_J^\mu R \pi_J^\mu\|_{L^2(\mu)} \leq 1$ we have $0 < \kappa_{J,i} \leq 1$.

Proposition 3.13. *For sufficiently large J it holds uniformly on Θ_s*

$$|\kappa_{J,1} - \kappa_1| + \|u_{J,1} - u_1\|_{H^1} \lesssim 2^{-Js}.$$

Proof. It suffices to show that $|\kappa_{J,1} - \kappa_1| + \|u_{J,1} - u_1\|_{L^2} \lesssim 2^{-J(s+1)}$. Indeed, by Jackson's and Bernstein's inequalities

$$\begin{aligned}
\|u_{J,1} - u_1\|_{H^1} &\leq \|u_{J,1} - \pi_J u_1\|_{H^1} + \|(I - \pi_J)\, u_1\|_{H^1} \lesssim 2^J \|u_{J,1} - \pi_J u_1\|_{L^2} + \|(I - \pi_J)\, u_1\|_{H^1} \\
&\lesssim 2^J \|u_{J,1} - u_1\|_{L^2} + 2^J \|(I - \pi_J) u_1\|_{L^2} + \|(I - \pi_J)\, u_1\|_{H^1} \\
&\lesssim 2^J \|u_{J,1} - u_1\|_{L^2} + 2^{-Js}
\end{aligned}$$

where we used the upper bound (3.18).

Recall that R is a compact self-adjoint positive-definite operator on $L^2(\mu)$. Furthermore

$$\begin{aligned}
\|(I - \pi_J^\mu)\, u_1\|_{L^2(\mu)} &\lesssim \|(I - \pi_J^\mu)\,(I - \pi_J)\, u_1\|_{L^2} \lesssim \|(I - \pi_J)\, u_1\|_{L^2} \\
&\lesssim 2^{-J(s+1)} \|u_1\|_{H^{s+1}} \lesssim 2^{-J(s+1)}.
\end{aligned}$$

Consequently, since by Lemma 3.5 operator R has a uniform spectral gap inequality

$$\|(I - \pi_J^\mu)\, u_1\|_{L^2(\mu)} \leq \frac{\kappa_1 - \kappa_2}{4\kappa_1}$$

holds for J large enough. It follows that we can use Theorem A.1 obtaining

$$|\kappa_{J,1} - \kappa_1| + \left\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \right\|_{L^2(\mu)} \lesssim 2^{-J(s+1)}.$$

The claim follows since $\|u_{J,1} - u_1\|_{L^2} \lesssim \left\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \right\|_{L^2(\mu)}$ by the equivalence of norms $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^2(\mu)}$. □

Corollary 3.14. *Projected operators $\pi_J^\mu R \pi_J^\mu$ have a uniform spectral gap, i.e. there exists $s_1 > 0$ such that*

$$\min\{|\kappa_{J,1}|, |\kappa_{J,2} - \kappa_{J,1}|\} \geq s_1$$

for every J large enough.

Proof. Follows from the proof of Theorem A.1. □

3.5.4. Analysis of the stochastic error

Define the operator $R_J : V_J \to V_J$ as the restriction of the operator $\pi_J^\mu R \pi_J^\mu$ to the finite dimensional Hilbert space V_J. Recall that the operator G_J was defined by the Gram matrix of the inner product $\langle \cdot, \cdot \rangle_\mu$, i.e. for $v \in V_J$ we have $\langle v, G_J v \rangle = \langle v, v \rangle_\mu$. Note that by (3.20)

$$R_J u_{J,i} = \kappa_{J,i} G_J u_{J,i}, \tag{3.21}$$

hence $(\kappa_{J,i}, u_{J,i})$ are solutions of generalized symmetric eigenvalue problem for R_J, G_J. When matrix $\widehat{G}_J$ is invertible the corresponding generalized eigenvalue problem for $\widehat{G}_J, \widehat{R}_J$, namely

$$\widehat{R}_J \widehat{u}_{J,i} = \widehat{\kappa}_{J,i} \widehat{G}_J \widehat{u}_{J,i} \tag{3.22}$$

has $\dim V_J$ solutions that we denote by $(\widehat{\kappa}_{J,i}, \widehat{u}_{J,i})$, $i = 0, 1, ..., \dim V_J - 1$. Recall that the eigenfunctions $\widehat{u}_{J,i}$ are normalized in $L^2[0,1]$.

In this subsection we want to bound the expected error between $(\kappa_{J,1}, u_{J,1})$ and $(\widehat{\kappa}_{J,1}, \widehat{u}_{J,1})$. From the general theory of a posteriori error bound techniques for generalized symmetric eigenvalue problems (see Section A.3) we know that the error between the eigenpairs can be controlled by the norm of the residual vectors:

$$r = (\widehat{R}_J - R_J) u_{J,1} + \kappa_{J,1}(G_J - \widehat{G}_J) u_{J,1} \text{ or } r^* = (R_J - \widehat{R}_J)\widehat{u}_{J,1} + \widehat{\kappa}_{J,1}(\widehat{G}_J - G_J)\widehat{u}_{J,1}.$$

Since the eigenpair $(\widehat{\kappa}_{J,1}, \widehat{u}_{J,1})$ of the problem (3.22) is random and depends on operators $\widehat{R}_J$ and $\widehat{G}_J$ it is easier to analyze the norm of the vector r rather than r^* (cf. Lemmas 3.15 and 3.16 where v is a deterministic function). Consequently in the following we refer to r as the residual vector. In the notation of Section A.3 we treat the deterministic problem (3.21) as a perturbed approximation of the data dependent problem (3.22).

Lemma 3.15. *For any $v \in V_J$ we have, uniformly on $\Theta_s \times \Gamma$,*

$$\mathbb{E}_{\sigma,b,\gamma}\Big[\|(\widehat{G}_J - G_J)v\|_{L^2}^2\Big] \lesssim N^{-1} 2^J \|v\|_{L^2}^2.$$

Proof. Given Lemma 3.11, the proof is a straightforward estimate analogously to [44, Lemma 4.8]. □

Now, we are ready to prove Lemma 3.7:

Proof of Lemma 3.7. A standard Neumann series argument shows that $\widehat{G}_J$ is invertible on $\mathcal{T}_1$ with $\|\widehat{G}_J{}^{-1}\|_{L^2} \le 2\|G_J^{-1}\|_{L^2}$. Since the invariant density μ has a positive lower bound uniformly on Θ_s, for any $v \in V_J$ we have

$$\langle v, G_J v \rangle = \langle v, v \rangle_\mu = \|v\|_{L^2(\mu)}^2 \gtrsim \|v\|_{L^2}^2.$$

Hence the smallest eigenvalue of the operator G_J is uniformly separated from zero. This implies that G_J^{-1} is uniformly bounded in the operator norm. The classical Hilbert-Schmidt norm inequality yields

$$\|\widehat{G}_J - G_J\|_{L^2}^2 \le \sum_{|\lambda| \le J} \|(\widehat{G}_J - G_J)\psi_\lambda\|_{L^2}^2.$$

Consequently, by Lemma 3.15, $\mathbb{E}_{\sigma,b,\gamma}[\|\widehat{G}_J - G_J\|_{L^2}^2] \lesssim N^{-1}2^{2J}$ and $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_1) \leq N^{-1}2^{2J}$ follows from Chebyshev's inequality. $\square$

Lemma 3.16. *For any $v \in V_J$ we have, uniformly on $\Theta_s \times \Gamma$,*

$$\mathbb{E}_{\sigma,b,\gamma}\left[\|(\widehat{R}_J - R_J)v\|_{L^2}^2\right] \lesssim N^{-1}2^J\|v\|_{L^2}^2.$$

Proof. By Lemma 3.11 we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b,\gamma}\left[\|(\widehat{R}_J - R_J)v\|_{L^2}^2\right] &= \sum_{|\lambda|\leq J} \mathrm{Var}_{\sigma,b,\gamma}\left[\frac{1}{N}\sum_{n=0}^{N-1}\psi_\lambda(X_{\tau_n})v(X_{\tau_n})\right] \\
&\lesssim \sum_{|\lambda|\leq J} N^{-1}\mathbb{E}_{\sigma,b,\gamma}\left[\psi_\lambda^2(X_{\tau_1})v^2(X_0)\right] \\
&\lesssim N^{-1}\|\sum_{|\lambda|\leq J}\psi_\lambda^2\|_\infty \mathbb{E}_{\sigma,b,\gamma}[v^2(X_0)] \\
&\lesssim N^{-1}2^J\|v^2\|_{L^2(\mu)}^2.
\end{aligned}$$

$\square$

Corollary 3.17. *We have, uniformly on $\Theta_s \times \Gamma$, the following bound on the norm of the residual vector $r = (\widehat{R}_J - R_J)u_{J,1} + \kappa_{J,1}(G_J - \widehat{G}_J)u_{J,1}$*

$$\mathbb{E}_{\sigma,b,\gamma}\left[\|r\|_{L^2}^2\right] \lesssim N^{-1}2^J.$$

Proof. Note that from Proposition 3.13 we know that, for J big enough, the eigenvalue $\kappa_{J,1}$ is uniformly bounded. Consequently

$$\mathbb{E}_{\sigma,b,\gamma}\left[\|r\|_{L^2}^2\right] \lesssim \mathbb{E}_{\sigma,b,\gamma}\left[\|(\widehat{R}_J - R_J)u_1^J\|_{L^2}^2\right] + \mathbb{E}_{\sigma,b,\gamma}\left[\|(\widehat{G}_J - G_J)u_1^J\|_{L^2}^2\right] \lesssim N^{-1}2^J$$

by Lemmas 3.15 and 3.16. $\square$

Proposition 3.18. *On the event $\mathcal{T}_1$ the eigenpair $(\widehat{\kappa}_{J,1}, \widehat{u}_{J,1})$ is the biggest nontrivial eigenpair of the matrix $\widehat{G}_J^{-1}\widehat{R}_J$. Furthermore there exists a set $\mathcal{T}_2 \subset \mathcal{T}_1$ such that*

$$\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1}2^{3J}$$

and

$$\mathbb{E}_{\sigma,b,\gamma}\left[\mathbf{1}_{\mathcal{T}_2}\cdot\left(|\kappa_{J,1} - \widehat{\kappa}_{J,1}|^2 + \|u_{J,1} - \widehat{u}_{J,1}\|_{L^2}^2\right)\right] \lesssim N^{-1}2^J$$

holds uniformly on Θ_s.

Proof. By Theorem A.10 there exists some $0 \leq i_0 \leq \dim V_J - 1$ such that the eigenpair $(\widehat{\kappa}_{J,i_0}, \widehat{u}_{J,i_0})$ of the problem (3.22) satisfies

$$\begin{aligned}
|\kappa_{J,1} - \widehat{\kappa}_{J,i_0}| &\leq \|\widehat{G}_J^{-1}\|_{L^2}\,\|r\|_{L^2}\,, \\
\|u_{J,1} - \widehat{u}_{J,i_0}\|_{L^2} &\leq \frac{2\sqrt{2}}{\delta(\widehat{\kappa}_{J,i})}\|\widehat{G}_J\|_{L^2}^{1/2}\|\widehat{G}_J^{-1}\|_{L^2}^{3/2}\,\|r\|_{L^2}\,,
\end{aligned}$$

where $\delta(\widehat{\kappa}_{J,i_0}) = \min_{j\neq i_0}\{|\widehat{\kappa}_{J,j} - \kappa_{J,1}|\}$ is the isolation distance of the eigenvalues $\widehat{\kappa}_{J,i_0}$ and $\kappa_{J,1}$. Let s_1 be the uniform spectral gap of operators R_J (see Corollary 3.14). Define $\mathcal{T}_2$ as the subset of $\mathcal{T}_1$ for which $i_0 = 1$ and $\delta(\widehat{\kappa}_{J,1}) \geq \frac{1}{2}s_1$. Since $\|\widehat{G}_J^{-1}\|_{L^2}$ and $\|\widehat{G}_J\|_{L^2}$ are uniformly bounded on the event $\mathcal{T}_1$ and $\mathbb{E}_{\sigma,b,\gamma}[\|r\|_{L^2}^2] \lesssim N^{-1}2^J$ the desired error bound holds when we restrict to the event $\mathcal{T}_2$.

To finish the proof we must show that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1}2^{3J}$. Denote

$$\mathcal{T}_2 = \mathcal{T}_1 \cap \underbrace{\{i_0 = 1\}}_{\mathcal{T}_{2,1}} \cap \underbrace{\{\delta(\widehat{\kappa}_{J,1}) \geq s_1/2\}}_{\mathcal{T}_{2,2}}.$$

First, using the absolute Weyl theorem (Theorem A.11) we observe that for any $0 \leq j \leq \dim V_J - 1$

$$\begin{aligned}
\mathbb{E}_{\sigma,b,\gamma}\left[\mathbf{1}_{\mathcal{T}_1} \cdot |\kappa_{J,j} - \widehat{\kappa}_{J,j}|^2\right] &\leq \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1} \cdot \|\widehat{G}_J^{-1}\|_{L^2}^2 \|(R_J - \widehat{R}_J) - \kappa_{J,j}(G_J - \widehat{G}_J)\|_{L^2}^2\Big] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1} \cdot \|R_J - \widehat{R}_J\|_{L^2}^2\Big] + \kappa_{J,j}\mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_0} \cdot \|G_J - \widehat{G}_J\|_{L^2}^2\Big] \\
&\lesssim N^{-1}2^{2J}
\end{aligned}$$

by the classical Hilbert-Schmidt norm inequality. Consequently, using the uniform lower bound on the spectral gap of R_J, we obtain

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \setminus \mathcal{T}_{2,1}) &\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot \big|\kappa_{J,2} - \kappa_{J,1}\big|^2\Big] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot \big|\kappa_{J,i_0} - \kappa_{J,1}\big|^2\Big] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot \big|\kappa_{J,i_0} - \widehat{\kappa}_{J,i_0}\big|^2\Big] + \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot \big|\widehat{\kappa}_{J,i_0} - \kappa_{J,1}\big|^2\Big] \\
&\lesssim N^{-1}2^{2J}.
\end{aligned}$$

Consider now the event $\mathcal{T}_{2,2}$. Since

$$\begin{aligned}
\delta(\widehat{\kappa}_{J,1}) &= \min_{j \neq 1} |\widehat{\kappa}_{J,j} - \kappa_{J,1}| \geq \min_{j \neq 1} \{|\kappa_{J,j} - \kappa_{J,1}| - |\widehat{\kappa}_{J,j} - \kappa_{J,j}|\} \\
&\geq s_1 - \max_{j \neq 1} \{|\widehat{\kappa}_{J,j} - \kappa_{J,j}|\},
\end{aligned}$$

we have

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \setminus \mathcal{T}_{2,2}) &\leq \mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \cap \{\max_{j \neq 1} \{|\widehat{\kappa}_{J,j} - \kappa_{J,j}|\} \geq s_1/2\}) \\
&\leq \sum_{1 < j \leq \dim V_J - 1} \mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \cap \{|\widehat{\kappa}_{J,j} - \kappa_{J,j}| \geq s_1/2\}) \\
&\lesssim \sum_{1 < j \leq \dim V_J - 1} \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_1} \cdot \big|\widehat{\kappa}_{J,j} - \kappa_{J,j}\big|^2\Big] \lesssim N^{-1}2^{3J}.
\end{aligned}$$

□

3.5.5. Proof of Theorem 3.8

From now on we choose $2^J \sim N^{1/(2s+3)}$. Recall that the biggest negative eigenvalue of the infinitesimal generator L is denoted by v_1 which is estimated by $\widehat{v}_{J,1}$ from (3.10).

Lemma 3.19. *Choose $2^J \sim N^{1/(2s+3)}$. There is an event $\mathcal{T}_3 \subset \mathcal{T}_2$ satisfying $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_3) \lesssim$*

$N^{-2s/(2s+3)}$ uniformly on $\Theta_s \times \Gamma$ and

$$\sup_{(\sigma,b\gamma)\in\Theta_s\times\Gamma} \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_3}|v_1-\widehat{v}_{J,1}|^2] \lesssim N^{-\frac{2s}{2s+3}}.$$

In particular, we can assume that $|\widehat{v}_{J,1}|$ is uniformly bounded on $\mathcal{T}_3$.

Proof. For convenience we denote $m := \min I$, $M := \max I$. On $\mathcal{T}_2$ we have $\widehat{\kappa}_{J,1} > 0$ and thus $\widehat{\kappa}_{J,1} = \widehat{\mathcal{L}}(-\widehat{v}_{J,1})$.

Step 1: Let us start with a consistency result for $\widehat{v}_{J,1}$. Since $\widehat{\mathcal{L}}$ is non-increasing and continuous, we have for any fixed $\varepsilon \in (0, C_1)$ with C_1 from (3.17) that

$$\mathbb{P}_\gamma(|\widehat{v}_{J,1}-v_1|<\varepsilon) \geq \mathbb{P}_\gamma\big(\widehat{\mathcal{L}}(-v_1+\varepsilon) < \widehat{\kappa}_{J,1} < \widehat{\mathcal{L}}(-v_1-\varepsilon)\big).$$

Using

$$\delta := \alpha m e^{(v_1-\varepsilon)M} \leq \inf_{\gamma\in\Gamma}\inf_{|y+v_1|\leq\varepsilon} |\mathcal{L}'_\gamma(y)|, \tag{3.23}$$

we have $|\mathcal{L}_\gamma(-v_1)-\mathcal{L}_\gamma(-v_1\pm\varepsilon)| \geq \delta\varepsilon$ uniformly in $\gamma\in\Gamma$ and

$$\begin{aligned}
&\mathbb{P}_{\sigma,b,\gamma}(|\widehat{v}_{J,1}-v_1|\geq\varepsilon)\\
&\quad\leq \mathbb{P}_{\sigma,b,\gamma}\big(\kappa_1-\widehat{\kappa}_{J,1} > \kappa_1-\widehat{\mathcal{L}}(-v_1+\varepsilon)\big) + \mathbb{P}_{\sigma,b,\gamma}\big(\widehat{\kappa}_{J,1}-\kappa_1 > \widehat{\mathcal{L}}(-v_1-\varepsilon)-\kappa_1\big)\\
&\quad\leq \sum_{y\in\{-\varepsilon,+\varepsilon\}} \mathbb{P}_{\sigma,b,\gamma}\big(|\widehat{\kappa}_{J,1}-\kappa_1| + |\widehat{\mathcal{L}}(-v_1+y)-\mathcal{L}_\gamma(-v_1+y)| > \delta\varepsilon\big)\\
&\quad\leq 2\mathbb{P}_{\sigma,b}\big(|\widehat{\kappa}_{J,1}-\kappa_1| > \frac{\delta\varepsilon}{2}\big) + \sum_{y\in\{-\varepsilon,+\varepsilon\}} \mathbb{P}_\gamma\big(|\widehat{\mathcal{L}}(-v_1+y)-\mathcal{L}_\gamma(-v_1+y)| > \frac{\delta\varepsilon}{2}\big).
\end{aligned}$$

By Propositions 3.13 and 3.18 and Markov's inequality the first probability is of the order $N^{-2s/(2s+3)}$ if $2^J \sim N^{1/(2s+3)}$. For the estimation error of $\widehat{\mathcal{L}}$ Markov's inequality yields for any $y>0$

$$\begin{aligned}
\mathbb{P}_\gamma\big(|\widehat{\mathcal{L}}(y)-\mathcal{L}_\gamma(y)| > \delta\varepsilon/2\big) &\leq 2(\delta\varepsilon)^{-2}\mathbb{E}_\gamma\big[|\widehat{\mathcal{L}}(y)-\mathcal{L}_\gamma(y)|^2\big]\\
&= \frac{2}{N\delta^2\varepsilon^2}\operatorname{Var}_\gamma\big(e^{-y\Delta_1}\big) \leq \frac{2\mathcal{L}_\gamma(2y)}{N\delta^2\varepsilon^2}.
\end{aligned}$$

Therefore,

$$\mathbb{P}_{\sigma,b,\gamma}(|\widehat{v}_{J,1}-v_1|\geq\varepsilon) \lesssim N^{-2s/(2s+3)}. \tag{3.24}$$

Step 2: To determine the rate of $\widehat{v}_{J,1}$, we use a Taylor expansion which yields for some intermediate point ξ between $-v_1$ and $-\widehat{v}_{J,1}$

$$\widehat{\kappa}_{J,1} = \widehat{\mathcal{L}}(-\widehat{v}_{J,1}) = \widehat{\mathcal{L}}(-v_1) + (v_1-\widehat{v}_{J,1})\widehat{\mathcal{L}}'(\xi).$$

Since on the other hand we have $\widehat{\kappa}_{J,1} = \mathcal{L}_\gamma(-v_1) + \widehat{\kappa}_{J,1} - \kappa_1$, we conclude

$$v_1 - \widehat{v}_{J,1} = \frac{\mathcal{L}_\gamma(-v_1) - \widehat{\mathcal{L}}(-v_1) + \widehat{\kappa}_{J,1} - \kappa_1}{\widehat{\mathcal{L}}'(\xi)},$$

provided the denominator can be uniformly bounded with high probability. By (3.24) the event $\mathcal{T}_{3,1} := \{|\widehat{v}_{J,1}-v_1|<\varepsilon\}$ has at least the probability $1-cN^{-2s/(2s+3)}$ for some $c>0$.

On $\mathcal{T}_{3,1}$ we have

$$|\widehat{\mathcal{L}}'(\xi)| \geq \inf_{|y+v_1|<\varepsilon} \mathcal{L}'_\gamma(y) - \sup_{|y+v_1|<\varepsilon} |\widehat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)|.$$

With δ from (3.23) we conclude that $|\widehat{\mathcal{L}}'(\xi)| \geq \delta/2$ on the event

$$\mathcal{T}_{3,2} := \{ \sup_{y\in[-v_1-\varepsilon,-v_1+\varepsilon]} \big|\widehat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)\big|^2 < \delta/2 \}.$$

Note that in $\mathcal{T}_{3,2}$ we take the supremum of the empirical processes related to $(\Delta_n)_{n=1,\dots,N}$ acting on the function set $\mathcal{F} := \{[0,\infty) \ni x \mapsto xe^{-yx} : y \in [|v_1| - \varepsilon, |v_1| + \varepsilon]\}$. Since $\mathcal{F}$ is the multiplication of the identity map with the transition class $\{e^{-yx} : y > 0\})$, $\mathcal{F}$ is a Vapnik-Červonenkis class and admits the constant envelope function $(|v_1| - \varepsilon)^{-1} e^{-1}$. The empirical process theory (e.g., van der Vaart and Wellner [92], Thm. 2.14.1) yields

$$\mathbb{E}_\gamma\Big[\sup_{y\in[-v_1-\varepsilon,-v_1+\varepsilon]} \big|\widehat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)\big|^2\Big] \lesssim \frac{1.}{N(|v_1| - \varepsilon)^2}$$

and by Markov's inequality $\mathbb{P}_\gamma(\Omega \setminus \mathcal{T}_{3,2}) \lesssim 1/N$. With $\mathcal{T}_3 := \mathcal{T}_{3,1} \cap \mathcal{T}_{3,2} \cap \mathcal{T}_2$ we finally obtain

$$\begin{aligned}\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_3}|v_1 - \widehat{v}_{J,1}|^2] &\leq 2\mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_3}\frac{|\mathcal{L}_\gamma(-v_1) - \widehat{\mathcal{L}}(-v_1)|^2 + |\bar{\kappa}_1 - \kappa_1|^2}{|\widehat{\mathcal{L}}'(\xi)|^2}\Big] \\ &\lesssim N^{-1} + \mathbb{E}_{\sigma,b,\gamma}\big[\mathbf{1}_{\mathcal{T}_3}|\widehat{\kappa}_{J,1} - \kappa_1|^2\big] \lesssim N^{-2s/(2s+3)}.\end{aligned}$$

□

Corollary 3.20. *Choosing $2^J \sim N^{1/(2s+3)}$, there exists an event $\mathcal{T}_4 = \mathcal{T}_0 \cap \mathcal{T}_3$ of high probability, i.e. $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_4) \lesssim N^{-2s/(2s+3)}$, such that the estimators $\widehat{\mu}_J$ and $\widehat{v}_{J,1}$ are uniformly bounded on $\mathcal{T}_4$. Furthermore, for N big enough, we have uniformly on Θ_s and Γ*

$$\mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_4} \cdot \Big(\big|v_1 - \widehat{v}_{J,1}\big|^2 + \|u_1 - \widehat{u}_{J,1}\|^2_{H^1}\Big)\Big] \quad \lesssim \quad N^{-2s/(2s+3)}.$$

Proof. Note that $\mathcal{T}_4$ is a subset of the events from Proposition 3.18, Lemma 3.19 and the event that $\widehat{\mu}_J$ is uniformly bounded from below and above (see Proposition 3.12). Then $\mathcal{T}_4$ is a high probability event and by Propositions 3.13 and 3.18, the choice $2^J \sim N^{1/(2s+3)}$ yields the claimed bound of the expectation. □

Before we present the proof of Theorem 3.8 we need to another representation of the volatility estimator which allows us to bound the derivative of the estimated eigenfunction.

Lemma 3.21. *Set $0 < a < b < 1$. There exists a high probability event $\mathcal{T}_5 \subset \mathcal{T}_4$ satisfying $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-2s/(2s+3)}$ and such that*

$$\mathbf{1}_{\mathcal{T}_5} \cdot \widehat{\sigma}^2_J(x) = \mathbf{1}_{\mathcal{T}_5} \cdot \frac{2\widehat{v}_{J,1} \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{(\widehat{u}'_{J,1}(x) \vee c'_{a,b})\widehat{\mu}_J(x)} \wedge D,$$

for a deterministic constant $c'_{a,b} > 0$ satisfying $c'_{a,b} \leq c_{a,b} \leq \inf_{x\in[a,b]} u'_1(x)$.

Proof. Recall that

$$\widehat{\sigma}^2_J(x) = \frac{2\widehat{v}_{J,1} \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widehat{u}'_{J,1}(x)\widehat{\mu}_J(x)} \wedge D = \frac{2\widehat{v}_{J,1} \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widehat{\mu}_J(x)\big(\widehat{u}'_{J,1}(x) \vee \frac{2\widehat{v}_{J,1} \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widehat{\mu}_J(x)D}\big)}.$$

Let $m = \frac{1}{2}\inf \mu(x)$ and $M = 2\sup \widehat{\mu}_J$. By Proposition 3.12 $m \le \widehat{\mu}_J(x) \le M$ for all $x \in [0,1]$ on the event $\mathcal{T}_0$. This event is especially contained in

$$\mathcal{T}_5 := \mathcal{T}_4 \cap \Big\{4\|\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy - v_1\int_0^x u_1(y)\mu(y)dy\|_\infty \le d^2 c_{a,b} m\Big\},$$

where $\mathcal{T}_4$ is the high probability event from Corollary 3.20. On $\mathcal{T}_5$ it holds

$$\begin{aligned}
&\frac{2\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{D\widehat{\mu}_J(x)} \\
&\qquad \ge \frac{2v_1\int_0^x u_1(y)\mu(y)dy - 2|\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy - v_1\int_0^x u_1(y)\mu(y)dy|}{D\widehat{\mu}_J(x)} \\
&\qquad = \frac{\sigma^2(x)u_1'(x)\mu(x) - 2|\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy - v_1\int_0^x u_1(y)\mu(y)dy|}{D\widehat{\mu}_J(x)} \\
&\qquad \ge \frac{d^2 c_{a,b} m}{2MD} =: c'_{a,b}.
\end{aligned}$$

Furthermore, by Corollary 3.20, using Markov and triangle inequalities, it is easy to check that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-\frac{2s}{2s+3}}$, cf. estimate (3.25). □

Proof for the volatility estimator. Set $0 < a < b < 1$. Note first that since $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-\frac{2s}{2s+3}}$ and $\sigma, \widehat{\sigma}$ are bounded we just have to verify that $\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_5}\cdot\|\sigma^2 - \widehat{\sigma}^2\|^2_{L^2([a,b])}] \lesssim N^{-\frac{2s}{2s+3}}$. Denote $\widetilde{u}'_{J,1}(x) = \widehat{u}'_{J,1}(x) \vee c'_{a,b}$ and $\widetilde{\sigma}^2_J(x) = \frac{2\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widetilde{u}'_{J,1}(x)\widehat{\mu}_J(x)}$. Since for $x \in [a,b]$ the functions u_1' and μ are uniformly separated from zero, we have that on $\mathcal{T}_5$

$$\begin{aligned}
&|\sigma^2(x) - \widetilde{\sigma}^2_J(x)| \\
&\quad \le \Big|\frac{2v_1\int_0^x u_1(y)\mu(y)dy}{u_1'(x)\mu(x)} - \frac{2\widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy}{\widetilde{u}'_{J,1}(x)\widehat{\mu}_J(x)}\Big| \\
&\quad = \Big|\frac{2\left(v_1\int_0^x u_1(y)\mu(y)dy - \widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy\right)}{u_1'(x)\mu(x)} - \frac{\widetilde{\sigma}^2_J(x)\left(u_1'(x)\mu(x) - \widetilde{u}'_{J,1}(x)\widehat{\mu}_J(x)\right)}{u_1'(x)\mu(x)}\Big| \\
&\quad \lesssim \Big|v_1\int_0^x u_1(y)\mu(y)dy - \widehat{v}_{J,1}\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy\Big| + |\widetilde{\sigma}^2_J(x)|\Big|\frac{u_1'(x)\mu(x) - \widetilde{u}'_{J,1}(x)\widehat{\mu}_J(x)}{u_1'(x)}\Big| \\
&\quad =: A_1(x) + A_2(x).
\end{aligned}$$

Observe that since $\widehat{\mu}_J$ is uniformly bounded on the event $\mathcal{T}_5$ and since the eigenfunction $\widehat{u}_1$ is normalized, the Cauchy-Schwarz inequality grants that $\int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy$ is uniformly bounded. Hence,

$$\begin{aligned}
A_1(x) &= |v_1(\int_0^x u_1(y)\mu(y)dy - \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy) + \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy(v_1 - \widehat{v}_{J,1})| \\
&\lesssim |\int_0^x u_1(y)\mu(y)dy - \int_0^x \widehat{u}_{J,1}(y)\widehat{\mu}_J(y)dy| + |v_1 - \widehat{v}_{J,1}| \\
&\lesssim |\int_0^x u_1(y)(\mu(y) - \widehat{\mu}_{J,1}(y))dy| + |\int_0^x (u_1(y) - \widehat{u}_{J,1}(y))\widehat{\mu}_J(y)dy| + |v_1 - \widehat{v}_{1,J}| \\
&\le \|u_1\|_{L^2}\|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{L^2}\|\widehat{\mu}_J\|_{L^2} + |v_1 - \widehat{v}_{1,J}|
\end{aligned}$$

$$= \|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{L^2} + |v_1 - \widehat{v}_{J,1}|. \tag{3.25}$$

Furthermore, since $\widetilde{\sigma}_J^2(x)$ is uniformly bounded on $\mathcal{T}_5$

$$\begin{aligned} A_2(x) &\lesssim |\mu(x) - \widehat{\mu}_J(x)| + \frac{|\widehat{\mu}_J(x)|}{|u_1'(x)|}|u_1'(x) - \widetilde{u}_{J,1}'(x)| \\ &\lesssim |\mu(x) - \widehat{\mu}_J(x)| + |u_1'(x) - \widetilde{u}_{J,1}'(x)| \\ &\lesssim |\mu(x) - \widehat{\mu}_J(x)| + |u_1'(x) - \widehat{u}_{J,1}'(x)|. \end{aligned} \tag{3.26}$$

Consequently,

$$\begin{aligned} \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_5} \cdot \|\sigma^2 - \widehat{\sigma}_J^2\|_{L^2}^2\Big] &\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_5} \cdot (\|A_1\|_{L^2}^2 + \|A_2\|_{L^2}^2)\Big] \\ &\lesssim \mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_5} \cdot \Big(\|\mu - \widehat{\mu}_J\|_{L^2}^2 + \|u_1 - \widehat{u}_{J,1}\|_{H^1}^2 + |v_1 - \widehat{v}_{J,1}|^2\Big)\Big] \\ &\lesssim N^{-2s/(2s+3)}. \end{aligned}$$

□

Proof for the drift estimator. To obtain the upper bound on the drift term first note that using Bernstein's inequality we can extend the proofs of Propositions 3.13 and 3.18 to obtain

$$\mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_4} \cdot \|u_1 - \widehat{u}\|_{H^2}^2\Big] \lesssim N^{-\frac{2(s-1)}{2s+3}}. \tag{3.27}$$

Let $\mathcal{T}_6 = \mathcal{T}_5 \cap \{\inf_{x\in[a,b]} \widehat{u}_{J,1}'(x) \geq c_{a,b}/2\} \cap \{\|\widehat{u}_{J,1}\|_{H^2} \leq 2\|u_1\|_{H^2}\}$. By Lemma 3.21 and (3.27) we obtain that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_6) \lesssim N^{-\frac{2(s-1)}{2s+1}}$. Since both b and $\widehat{b}$ are bounded in L^2, we can restrict the error analysis to the high probability event $\mathcal{T}_6$. Recall the definition of $\widetilde{b}$ from (3.12). Since $\|b\|_{L^2([a,b])} \leq D$ we have $\|\widehat{b}_J - b\|_{L^2([a,b])} \leq \|\widetilde{b}_J - b\|_{L^2([a,b])}$. Consequently, it remains to show

$$\mathbb{E}_{\sigma,b,\gamma}\Big[\mathbf{1}_{\mathcal{T}_6} \cdot \|\widetilde{b}_J - b\|_{L^2([a,b])}\Big] \lesssim N^{-\frac{2(s-1)}{2s+3}}.$$

On $\mathcal{T}_6$, for $x \in [a,b]$ we have

$$\begin{aligned} |\widetilde{b}_J(x) - b(x)| \leq& \Big|\frac{\widehat{v}_{J,1}\widehat{u}_{J,1}(x)}{\widehat{u}_{J,1}'(x)} - \frac{\widetilde{\sigma}_J^2(x)\widehat{u}_{J,1}''(x)}{2\widehat{u}_{J,1}'(x)} - \frac{v_1 u_1(x)}{u_1'(x)} + \frac{\sigma^2(x)u_1''(x)}{2u_1'(x)}\Big| \\ \leq& |u_1'(x)|^{-1}\Big|\widehat{v}_{J,1}\widehat{u}_{J,1}(x) - v_1u_1(x) + \frac{\sigma^2(x)}{2}u_1''(x) - \frac{\widetilde{\sigma}_J^2(x)}{2}\widehat{u}_{J,1}''(x)\Big| \\ &+ \frac{|\widetilde{b}_J(x)|}{|u_1'(x)|}|u_1'(x) - \widehat{u}_{J,1}'(x)|. \end{aligned}$$

The uniform lower bound on $|u_1'|$ yields

$$\begin{aligned} \|\widetilde{b}_J - b\|_{L^2([a,b])}^2 \lesssim& \|\widehat{v}_{J,1}\widehat{u}_{J,1} - v_1u_1\|_{L^2([a,b])}^2 + \|\widetilde{\sigma}_J^2\widehat{u}_{J,1}'' - \sigma^2 u_1''\|_{L^2([a,b])}^2 \\ &+ \|\widetilde{b}_J\|_{L^2([a,b])}^2\|\widehat{u}_{J,1}' - u_1'\|_{L^\infty([a,b])}^2 \\ =:& B_1 + B_2 + B_3. \end{aligned}$$

We will estimate these three terms separately. Corollary 3.20 and the normalization of $\widehat{u}_{J,1}$

yield

$$\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6}B_1] \le \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}}(|\widehat{v}_{J,1}-v_1|^2\|\widehat{u}_{J,1}\|_{L^2}^2+|v_1|^2\|\widehat{u}_{J,1}-u_1\|_{L^2}^2)] \lesssim N^{-2s/(2s+3)}.$$

The second term can be decomposed into

$$B_2 \le 2\|\widetilde{\sigma}_J^2-\sigma^2\|_\infty^2\|u_1''\|_{L^2}^2+2\|\widetilde{\sigma}_J^2\|_\infty^2\|\widehat{u}_{J,1}''-u_1''\|_{L^2}^2.$$

From (3.25) and (3.26) we can easily verify that

$$\|\widehat{\sigma}_J^2-\sigma^2\|_\infty \lesssim |\widehat{v}_{J,1}-v_1|+\|\widehat{u}_{J,1}-u_1\|_{H^2}+\|\widehat{\mu}_J-\mu\|_{H^1}.$$

Since $\widehat{\sigma}_J^2$ is bounded by construction, we conclude

$$\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6}B_2] \le \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6}(|\widehat{v}_{J,1}-v_1|^2+\|\widehat{u}_{J,1}-u_1\|_{H^2}^2+\|\widehat{\mu}_J-\mu\|_{H^1}^2)] \lesssim N^{-2(s-1)/(2s+3)}.$$

For the last term it holds

$$\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6}B_3] \le \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6}\|\widetilde{b}_J\|_{L^2([a,b])}^2\|\widehat{u}_{J,1}-u_1\|_{H^2}^2] \lesssim N^{-2(s-1)/(2s+3)}$$

since $\|\widetilde{b}_J\|_{L^2[(a,b)]}$ is uniformly bounded on $\mathcal{T}_6$. □

3.6. Proof of the lower bounds

First, note that estimating the sampling distribution γ has no impact on the convergence rates, because the Laplace transform can be estimated with the parametric rate. Therefore, it suffices to use the same distribution $\gamma \in \Gamma$ for all alternatives. Throughout this section we thus fix some $\gamma \in \Gamma$ which admits a bounded Lebesgue density on $[0,T]$ for some $T>0$.

Without loss of generality we can suppose that $(1,0)\in\Theta_s$. To construct the alternatives, let ψ be a compactly supported wavelet in H^s with one vanishing moment. We set $\psi_{jk}(x) = 2^{j/2}\psi(2^jx-k)$ and denote by $K_j\subset\mathbb{Z}$ a maximal set of indices k such that $\operatorname{supp}(\psi_{jk})\subset[a,b]$ and $\operatorname{supp}(\psi_{jk})\cap\operatorname{supp}(\psi_{jk'})=\emptyset$ holds for all $k,k'\in K_j$, $k\neq k'$. For a constant $\delta>0$ and all $\varepsilon=(\varepsilon_k)\in\{-1,1\}^{|K_j|}$ we define

$$S_\varepsilon(x)=S_\varepsilon(j,x)=\Big(2+\delta\sum_{k\in K_j}\varepsilon_k\psi_{jk}(x)\Big)^{-1}.$$

Choosing $\delta\sim 2^{-j(s+1/2)}$ yields $(\sqrt{2S_\varepsilon},S_\varepsilon')\in\Theta_s$. The corresponding diffusions $X^{(\varepsilon)}$ are defined by their generators

$$\begin{aligned} L_\varepsilon f(x) &= S_\varepsilon(x)f''(x)+S_\varepsilon'(x)f'(x),\\ \operatorname{dom}(L_\varepsilon) &= \operatorname{dom}(L). \end{aligned}$$

Note that for any ε the invariant measure of $X^{(\varepsilon)}$ is given by Lebesgue measure on $[0,1]$. For ε,ε' with $\|\varepsilon-\varepsilon'\|_{\ell^2}=2$ we have

$$S_{\varepsilon'}(x)-S_\varepsilon(x)=\pm2\delta\psi_{jk}(x)S_{\varepsilon'}(x)S_\varepsilon(x).$$

Since $S_\varepsilon, S_{\varepsilon'}$ converge uniformly to $1/2$ as $j \to \infty$, the L^2-distances of the volatility functions and the drift functions of the alternatives ε and ε' are bounded by

$$\|2S_{\varepsilon'} - 2S_\varepsilon\|_{L^2} \gtrsim \delta, \quad \|S'_{\varepsilon'} - S'_\varepsilon\|_{L^2} \gtrsim 2^j\delta.$$

Therefore, Assouad's lemma and $\delta \sim 2^{-j(s+1/2)}$ yield for all estimators $\bar{\sigma}^2$ and $\bar{b}$

$$\begin{aligned}\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b,\gamma}\Big[\|\bar{\sigma}^2 - \sigma^2\|^2_{L^2([a,b])}\Big] &\gtrsim 2^j\delta = 2^{-2sj},\\ \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b,\gamma}\Big[\|\bar{b} - b\|^2_{L^2([a,b])}\Big] &\gtrsim 2^{3j}\delta = 2^{-2(s+1)j},\end{aligned} \tag{3.28}$$

provided the Kullback-Leibler divergence between the distributions of $(X^{(\varepsilon)}_{\tau_n})_{n=0,\dots,N}$ and $(X^{(\varepsilon')}_{\tau_n})_{n=0,\dots,N}$ remains uniformly bounded for all alternatives $\varepsilon, \varepsilon'$ with $\|\varepsilon - \varepsilon'\|_{\ell^2} = 2$.

To bound the Kullback-Leibler divergence, we have to take into account the random observation times. Denote the transition density of $(X_t)_{t\geq 0}$ by $p_t(x,y)dy = \mathbb{P}_{\sigma,b}(X_t = dy|X_0 = x)$ for $x, y \in [0,1], t \geq 0$. By the independence of the observation time τ and the process X we have

$$Rf(x) = \mathbb{E}_{\sigma,b,\gamma}\left[f(X_\tau)|X_0 = x\right] = \int_0^\infty P_t f(x)\gamma(dt) = \int_0^\infty \int_0^1 p_t(x,y)f(y)dy\gamma(dt).$$

For one dimensional diffusions with bounded drift and differentiable, uniformly separated from zero, volatility we know that

$$p_t(x,y) \leq c_0\Big(1 + \frac{1}{\sqrt{t}}\Big)$$

with $c_0 > 0$ depending only on the bounds for the drift and volatility (see Qian and Zheng [71, Thm. 1]). The assumption $\mathbb{E}[\tau^{-1/2}] < \infty$ thus ensures that

$$r(x,y) = \int_0^\infty p_t(x,y)\gamma(dt)$$

is a well defined kernel of operator R. We obtain the following generalization of Proposition 6.4 in [44]:

Lemma 3.22. *Assume $\mathbb{E}_\gamma\left[\tau^{-1/2}\right] < \infty$. If $(\sigma_n, b_n) \in \Theta_s$, $n \geq 0$, such that*

$$\lim_{n\to\infty} \|\sigma_n - \sigma_0\|_\infty = 0 \quad \textit{and} \quad \lim_{n\to\infty} \|b_n - b_0\|_\infty = 0,$$

then the corresponding kernels $r^{(n)}(x,y)dy = \mathbb{P}_{\sigma_n,b_n}(X_\tau \in dy|X_0 = x)$ satisfy

$$\lim_{n\to\infty} \|r^{(n)} - r^{(0)}\|_\infty = 0.$$

Note that the bounded Lebesgue density γ near the origin specially ensures that $\mathbb{E}_\gamma[\tau^{-1/2}] < \infty$.

Proof. Due to the bound $\|p_t^{(n)}(\cdot,\cdot)\|_\infty \lesssim 1 + t^{-1/2}$, dominated convergence yields

$$\begin{aligned} \|r^{(n)} - r^{(0)}\|_\infty &= \sup_{x,y\in[0,1]} \Big| \int_0^\infty \Big(p_t^{(n)}(x,y) - p_t^{(0)}(x,y)\Big)\gamma(dt)\Big| \\ &\leq \int_0^\infty \|p_t^{(n)} - p_t^{(0)}\|_\infty \gamma(dt). \end{aligned}$$

By [44, Prop. 6.4] this tends to zero. □

Exactly as in [44, Sect. 5.2], this lemma allows us to bound the Kullback-Leibler divergence by $N\|r_{\varepsilon'} - r_\varepsilon\|^2_{L^2([0,1]^2)}$ for kernels $r_{\varepsilon'}$ and r_ε of $R_{\varepsilon'}$ and R_ε, respectively, for any $\varepsilon, \varepsilon'$ with $\|\varepsilon - \varepsilon'\|_{\ell^2} = 2$. Note that $\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0,1]^2)}$ is the Hilbert-Schmidt norm distance $\|R - R^{\varepsilon'}\|_{HS} = \|(R^\varepsilon - R^{\varepsilon'})|_V\|_{HS}$ where

$$V = \Big\{ f \in L^2([0,1]) \Big| \int_0^1 f = 0 \Big\}.$$

We will bound the Hilbert-Schmidt norm by the difference of the inverses of the generators, which are, in contrast to the generators itself, bounded operators. Recall that $R = \mathcal{L}(-L)$ for the Laplace transform $\mathcal{L}(z) = \int_0^\infty e^{-tz}\gamma(dt), z \geq 0$. By the functional calculus for operators the function $f(z) = \mathcal{L}\left(-z^{-1}\right)$ maps $(L_\varepsilon|_V)^{-1}$ to $R^\varepsilon|_V$. Furthermore, f is uniformly Lipschitz on $(-\infty, 0)$:

Lemma 3.23. *Suppose that $\gamma \in \Gamma$ admits a bounded Lebesgue density on $[0,T]$ for some $T > 0$. Then we have*

$$c := \sup_{z<0} \Big| \frac{1}{z^2} \int_0^\infty t e^{t/z} \gamma(dt) \Big| < \infty.$$

Proof. We decompose

$$\sup_{z<0} \Big| \frac{1}{z^2} \int_0^\infty t e^{t/z} \gamma(dt) \Big| \leq \sup_{z<0} \Big| \frac{1}{z^2} \int_0^T t e^{t/z} \gamma(dt) \Big| + \sup_{z<0} \Big| \frac{1}{z^2} \int_T^\infty t e^{t/z} \gamma(dt) \Big| =: S_1 + S_2.$$

Due to the bounded Lebesgue density on $[0,T]$, we estimate the first term by substituting $s = t/z$

$$S_1 \lesssim \sup_{z<0} z^{-2} \int_0^T t e^{t/z} dt = \sup_{z<0} \int_{T/z}^0 s e^s ds = \int_{-\infty}^0 s e^s ds < \infty.$$

For the second term note that the function $g_a(x) = x^2 e^{-ax}$ takes maximum at $x = 2/a$ and $g\left(2/a\right) = 4a^{-2}e^{-2}$. Consequently,

$$S_2 \leq \sup_{z<0} \int_T^\infty t g_t(|z|^{-1})\gamma(dt) = \int_T^\infty \frac{4}{te^2}\gamma(dt) \leq \frac{4}{Te^2} < \infty.$$ □

We conclude

$$\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0,1]^2)} = \big\|(R^\varepsilon - R^{\varepsilon'})|_V\big\|_{HS} \leq c\big\|(L_\varepsilon|_V)^{-1} - (L_{\varepsilon'}|_V)^{-1}\big\|_{HS} \lesssim \delta 2^{-j} = 2^{-j(2s+3)/2},$$

by the estimate for the difference of inverses of the generators that was established in [44, Sect. 5.3]. In order to bound $N\|r_{\varepsilon'} - r_\varepsilon\|^2_{L^2([0,1]^2)}$, we thus choose j such that $2^j \sim N^{1/(2s+3)}$. In view of (3.28) we have proven Theorem 3.9. □

3.7. Proof for the adaptive estimator

In order to show that Lepski's method works, we need the following concentration result. It slightly generalizes the corresponding concentration inequalities by Nickl and Söhl [65, Theorems 10 and 11] for a low-frequently observed reflected diffusion to random sampling times.

Proposition 3.24. *Grant Assumptions 3.1 and 3.3 with $s > 5/2$ and $\gamma \in \Gamma$, $\mathbb{E}_\gamma[\tau^{-1/2}] \le D$. There is a constant $c > 0$ depending only on d, D, I and α, such that, for any $\kappa > 0, N \in \mathbb{N}$ and any $f \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R}), g \in L^2(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$:*

$$\mathbb{P}_{\sigma,b,\gamma}\Big(\Big|\sum_{n=0}^{N}\big(f(X_{\tau_n}) - \mathbb{E}_{\sigma,b,\gamma}[f(X_0)]\big)\Big| > \kappa\Big) \lesssim \exp\Big(-c\min\Big\{\frac{\kappa^2}{N\|f\|^2_{L^2}}, \frac{\kappa}{(\log N)\|f\|_\infty}\Big\}\Big)$$

and

$$\mathbb{P}_{\sigma,b,\gamma}\Big(\Big|\sum_{n=0}^{N-1}\big(g(X_{\tau_n}, X_{\tau_{n+1}}) - \mathbb{E}_{\sigma,b,\gamma}[g(X_0, X_{\tau_1})]\big)\Big| > \kappa\Big)$$
$$\lesssim \exp\Big(-c\min\Big\{\frac{\kappa^2}{N\|g\|^2_{L^2}}, \frac{\kappa}{(\log N)\|g\|_\infty}\Big\}\Big).$$

Proof. The conditions of the Markov chain concentration result by Adamczak [2, Theorem 6] have to be verified. This can be done along the lines of the proofs in [65] using Lemma 3.5 and noting that the transition density of the time-changed chain $(X_{\tau_n})_{n\ge1}$ is given by $p_\gamma(x,y) = \int_0^\infty p_t(x,y)\gamma(dt)$ where $p_t(x,y)$ denotes the transition density of the diffusion $(X_t)_{t\ge0}$. The condition $s > 5/2$ ensures that the transition density p_γ is bounded from below uniformly on $[0,1]^2$. Indeed, $p_\gamma(x,y) \ge K\gamma(I) \ge K\alpha$, where K is the uniform lower bound on $\inf_{t\in I} p_t$ obtained in [65, Proposition 9]. Since $\|p_t\|_\infty \lesssim 1 + t^{-1/2}$, the condition $\mathbb{E}_\gamma[\tau^{-1/2}] < \infty$ ensures a uniform upper bound on p_γ. □

To analyze the performance of $\widetilde{\sigma}^2$, we first decompose its estimation error into a deterministic and a stochastic error term. In what follows, $C = C(d, D, I, \alpha)$ denotes a numeric constant which may vary from line to line. We deduce from the proof of Theorem 3.8 on the there defined event $\mathcal{T}_5$, that for any $J \in \mathcal{J}_N$

$$\begin{aligned}\|\widehat{\sigma}^2_J - \sigma^2\|_{L^2[a,b]} &\le C\big(\|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{H^1} + |v_1 - \widehat{v}_{J,1}|\big)\\ &\le C\big(\|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{H^1} + |\kappa_1 - \widehat{\kappa}_{J,1}| + |\mathcal{L}_\gamma(-v_1) - \widehat{\mathcal{L}}_\gamma(-v_1)|\big)\\ &\le D_J + S_J, \end{aligned} \tag{3.29}$$

where

$$\begin{aligned} D_J &:= C\big(\|(I - \pi_J)\mu\|_{L^2} + \|u_1 - u_{J,1}\|_{H^1} + |\kappa_1 - \kappa_{J,1}|\big),\\ S_J &:= C\big(\|\pi_J\mu - \widehat{\mu}_J\|_{L^2} + \|u_{J,1} - \widehat{u}_{J,1}\|_{H^1} + |\kappa_{J,1} - \widehat{\kappa}_{J,1}| + |\mathcal{L}_\gamma(-v_1) - \widehat{\mathcal{L}}_\gamma(-v_1)|\big).\end{aligned}$$

Due to the smoothness of the invariant measure, Jackson's inequality and Proposition 3.13, there is some $\beta > 0$, depending on ψ, d and D such that

$$D_J \leq \beta 2^{-Js}.$$

We need that S_J concentrates around zero. Recalling the definition of the residual vector

$$r = (\widehat{R}_J - R_J)u_{J,1} + \kappa_{J,1}(G_J - \widehat{G}_J)u_{J,1},$$

Bernstein's inequality and Theorem A.10 on generalized symmetric eigenvalue problems yield, on the event $\mathcal{T}_2$ from Proposition 3.18, that

$$\|u_{J,1} - \widehat{u}_{J,1}\|_{H^1} + |\kappa_{J,1} - \widehat{\kappa}_{J,1}| \leq C2^J \|u_{J,1} - \widehat{u}_{J,1}\|_{L^2} + |\kappa_{J,1} - \widehat{\kappa}_{J,1}| \leq \|r\|_{L^2}(C2^J + 1).$$

Corollary 3.25. *Under the conditions of Proposition 3.24, for any $\tau > 1$ there exist $\eta_1, \eta_2, \eta_3 > 1$, such that, for all J with $2^J \lesssim \frac{N}{(\log N)^2 \log\log N}$, we have*

$$\mathbb{P}_{\sigma,b,\gamma}\Big(\|\pi_J\mu - \widehat{\mu}_J\|_{L^2} > 2^{\frac{J}{2}}\eta_1\sqrt{\frac{\log\log N}{N}}\Big) \lesssim (\log N)^{-\tau}, \tag{3.30}$$

$$\mathbb{P}_{\sigma,b,\gamma}\Big(\|r\|_{L^2} > 2^{\frac{J}{2}}\eta_2\sqrt{\frac{\log\log N}{N}}\Big) \lesssim (\log N)^{-\tau}, \tag{3.31}$$

$$\mathbb{P}_{\sigma,b,\gamma}\Big(|\mathcal{L}_\gamma(-v_1) - \widehat{\mathcal{L}}_\gamma(-v_1)| > \eta_3\sqrt{\frac{\log\log N}{N}}\Big) \lesssim (\log N)^{-\tau}. \tag{3.32}$$

In particular, there is a $\Lambda > 0$ such that $\mathbb{P}_{\sigma,b,\gamma}(4S_J > s_J) \lesssim (\log N)^{-\tau}$ for $s_J = s_J(\Lambda)$ from (3.14).

Proof. Fix $\tau > 1$. Since $\|\psi_\lambda\|_\infty \lesssim 2^{|\lambda|/2}$, for $|\lambda| \leq J$, using Proposition 3.24 we obtain

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}\Big(|\langle\psi_\lambda, \mu - \mu_N\rangle| > \eta_1\sqrt{\frac{\log\log N}{N}}\Big) &\lesssim \exp\Big(-c\min\Big\{\frac{\eta_1^2 N(\log\log N)}{N\|\psi_\lambda\|_{L^2}^2}, \frac{\eta_1\sqrt{N(\log\log N)}}{(\log N)\|\psi_\lambda\|_\infty}\Big\}\Big)\\
&\lesssim \exp\Big(-c\eta_1\min\Big\{\log\log N, \frac{\sqrt{N(\log\log N)}}{(\log N)2^{J/2}}\Big\}\Big)\\
&\lesssim (\log N)^{-c\eta_1} \lesssim (\log N)^{-\tau},
\end{aligned}$$

for some η_1 big enough. Applying a usual chaining argument, this concentration inequality carries over to $\max_{|\lambda|\leq J}|\langle\psi_\lambda, \mu - \mu_N\rangle|$, cf. [11, Theorem 2.1] and [65, Theorem 12]. Since $\|\mu_J - \widehat{\mu}_J\|_{L^2}^2 = \sum_{|\lambda|\leq J}|\langle\psi_\lambda, \mu - \mu_N\rangle|^2$, it follows that

$$\begin{aligned}
&\mathbb{P}_{\sigma,b,\gamma}\Big(\|\pi_J\mu - \widehat{\mu}_J\|_{L^2}^2 > \eta_1^2 2^J \frac{\log\log N}{N}\Big)\\
&\qquad\lesssim \mathbb{P}_{\sigma,b,\gamma}\Big(\max_{|\lambda|\leq J}|\langle\psi_\lambda, \mu - \mu_N\rangle|^2 > \eta_1^2\frac{\log\log N}{N}\Big)\\
&\qquad\lesssim (\log N)^{-\tau}.
\end{aligned}$$

To prove (3.31), note first that since $|\kappa_{J,1}| \le 1$, we have

$$\|r\|_{L^2} \le \|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2} + \|(G_J - \widehat{G}_J)u_{J,1}\|_{L^2}.$$

By Proposition 3.13 $\|u_{J,1}\|_{L^2}, \|u_{J,1}\|_\infty \lesssim 1$ holds for J big enough. Using the second inequality in Proposition 3.24, we obtain

$$\begin{aligned}\mathbb{P}_{\sigma,b,\gamma}&\Big(|\langle\psi_\lambda, (\widehat{R}_J - R_J)u_{J,1}\rangle| > \eta_2\sqrt{\frac{\log\log N}{N}}\Big)\\ &\lesssim \exp\Big(-c\eta_2 \min\Big\{\frac{N(\log\log N)}{N}, \frac{\sqrt{N(\log\log N)}}{(\log N)2^{J/2}}\Big\}\Big) \lesssim (\log N)^{-C\eta_2} \lesssim (\log N)^{-\tau},\end{aligned}$$

for η_2 big enough. Since $\|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2} = \sum_{|\lambda|\le J}|\langle\psi_\lambda, (\widehat{R}_J - R_J)u_{J,1}\rangle|^2$, we conclude again that

$$\mathbb{P}_{\sigma,b,\gamma}\Big(\|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2} > \eta_2 2^{\frac{J}{2}}\sqrt{\frac{\log\log N}{N}}\Big) \lesssim (\log N)^{-\tau}.$$

Arguing similarly we deduce also $\mathbb{P}_{\sigma,b,\gamma}\Big(\|(G_J - \widehat{G}_J)u_{J,1}\|_{L^2} > \eta_2 2^{\frac{J}{2}}\sqrt{\frac{\log\log N}{N}}\Big) \lesssim (\log N)^{-\tau}$ and thus (3.31) holds.

The concentration inequality (3.32) follows from the classical Bernstein inequality. Indeed, we have

$$\widehat{\mathcal{L}}_\gamma(-v_1) - \mathcal{L}_\gamma(-v_1) = \frac{1}{N}\sum_{n=1}^N \xi_n \quad \text{with} \quad \xi_n := e^{v_1\Delta_n} - \mathbb{E}_\gamma[e^{v_1\Delta_n}],$$

where, by Assumption 3.1 the random variables ξ_n are independent, centered and deterministically bounded by 2 (because $v_1 < 0$). Since $\mathrm{Var}_\gamma(\xi_n) \le \mathcal{L}_\gamma(-2v_1) \le 1$, we can choose η_3 uniformly for all $\gamma \in \Gamma$. □

We can now prove the convergence rate for the adaptive estimator.

Proof of Theorem 3.10. Let us introduce the oracle projection level

$$J^* := \min\big\{J \in \mathcal{J}_N : \beta 2^{-Js} < s_J/4\big\}.$$

By the choice of $\mathcal{J}_N$ we deduce $2^{J^*} \sim (N/\log\log N)^{1/(2s+3)}$ and $s_{J^*}^2 \sim (\log\log N/N)^{2s/(2s+3)}$. Since the number of elements in $\mathcal{J}_N$ is of order $\log N$, Proposition 3.24 yields $\mathbb{P}_{\sigma,b,\gamma}(\mathcal{A}_N) \to 1$ for the event

$$\mathcal{A}_N := \big\{\forall J \in \mathcal{J}_N : 4S_J \le s_J\big\} \cap \mathcal{T}_6$$

with $\mathcal{T}_6$ from the proof of Theorem 3.8. Due to the decomposition (3.29), on $\mathcal{A}_N$ we have for every$J \in \mathcal{J}_N$:

$$\|\widehat{\sigma}_J^2 - \sigma^2\|_{L^2[a,b]} \le D_J + S_J \le \beta 2^{-Js} + s_J/4.$$

Hence, for all $J \ge J^*$, $J \in \mathcal{J}_N$, we obtain

$$\|\widehat{\sigma}_J^2 - \sigma^2\|_{L^2[a,b]} \le s_J/2,$$

and thus, by the triangle inequality,

$$\|\widehat{\sigma}_J^2 - \widehat{\sigma}_{J^*}^2\|_{L^2[a,b]} \le s_J,$$

for all $J \geq J^*$, $J \in \mathcal{J}_N$. By definition of $\widehat{J}$, we conclude that $\widehat{J} \leq J^*$ on the event $\mathcal{A}_N$. We conclude that

$$\|\widetilde{\sigma}^2 - \sigma^2\|_{L^2[a,b]} \leq \|\widehat{\sigma}^2_{\widehat{J}} - \widehat{\sigma}^2_{J^*}\|_{L^2[a,b]} + \|\widehat{\sigma}^2_{J^*} - \sigma^2\|_{L^2[a,b]} \leq s_{J^*} + \frac{1}{2} s_{J^*} \leq \frac{3}{2} s_{J^*}. \qquad \square$$

4. Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies

This chapter is an extended version of the paper Chorowski [21]. The nonparametric volatility estimation problem of a scalar diffusion process observed at equidistant time points is addressed. Using the spectral representation of the volatility in terms of the invariant density and an eigenpair of the infinitesimal generator the first known estimator that attains the minimax optimal convergence rates for both high and low-frequency observations is constructed. The proofs are based on a posteriori error bounds for generalized eigenvalue problems as well as the path properties of scalar diffusions and stochastic analysis. The finite sample performance is illustrated by a numerical example.

4.1. Introduction

Consider the problem of estimating the volatility of a diffusion process $(X_t, t \geq 0)$. The statistical properties depend, essentially, on the observation scheme. It is natural to assume discrete observations:

$$X_0, X_\Delta, ..., X_{N\Delta}, \quad \Delta > 0, \quad T = N\Delta.$$

The quality of an estimator is typically assessed by its asymptotic properties when the sample size N tends to infinity. The usual assumptions are either $\Delta \to 0$ or $T \to \infty$, which corresponds to high and low-frequency regimes, respectively. Different frequency assumptions require very different methods. Since the frequency regimes are a theoretical construct, for any given sample, we need to choose among high and low-frequency estimators. Therefore, it is of crucial interest to develop universal methods that will perform at optimal level regardless of the sampling frequency. In this chapter, the first nonparametric estimator of the volatility that attains minimax optimal rates in both high and low-frequency regimes is introduced. In the parametric setting, the problem of the universal scale estimation was first raised in Jacobsen [51, 52]. The constructed estimators were consistent and asymptotically Gaussian for all values of Δ, but nearly efficient for small values of Δ only. The estimation method, which relied on the use of the estimating functions, is different from the one applied in this chapter.

It is a well-known consequence of the Girsanov theorem that when T is fixed, the drift coefficient is not identifiable. Since we are interested in a universal scale method, we focus on the volatility estimation and, henceforth, treat drift as a nuisance parameter.

The existing high-frequency estimators (see Florens-Zmirou [38], Hoffmann [48], Jacod [53], Bandi and Phillips [9]) are based on the interpretation of the squared volatility as the instantaneous conditional variance of the process. Consequently, the assumption $\Delta \to 0$ is crucial for the consistency of these estimators, see [37] and [83, Section 3]. On the other hand, it has been conjectured that the minimax optimal low-frequency estimator introduced by Gobet, Hoffmann and Reiß (GHR) [44] also performs well in the high-frequency regime. This conjecture is based on the observation that the spectral representation of the volatility in terms of an eigenpair of the infinitesimal generator can be generalized by replacing the invariant density with the occupation density of the path $(X_t, t \leq T)$. While this generalization might be sufficient to obtain the consistency of the GHR estimator when applied to the high-frequency data, the numerical study reveals that the convergence rates are not optimal. The reason for this is that when the time horizon of the sample is fixed, the estimator inherits the poor regularity of the occupation density, which, contrary to the invariant density, is not linked to the regularity of the diffusion coefficients. As we show below, this difficulty can be solved with the appropriate averaging of the spectral estimator, which is the main motivation behind the Definition 4.6 of the universally optimal estimator. For more details, refer to Section 4.2.1.

Based on the spectral method, the low-frequency analysis of the universally optimal estimator is similar to [44, 23]. The real difficulty is in the high-frequency analysis, where the universal estimator is compared to the benchmark high-frequency estimator introduced by Florens-Zmirou [38] (see Section 4.2.2). In particular, we refer to the perturbation theory for bilinear coercive forms with Hölder regular coefficients, which is developed in Appendix A.

In the next sections, we present the construction of the frequency universal estimator and state the high and low-frequency convergence rates. In Section 4.2 we discuss the relation of the proposed estimator to the high and low-frequency benchmark estimators. Finite sample behaviour of the new estimator compared with the Florens-Zmirou and GHR estimators is illustrated in Section 4.2.3. In Section 4.2.4 we discuss the assumptions and possible extensions of the model. The proofs of the high and low-frequency convergence rates are shown in Sections 4.3 and 4.4, respectively.

4.1.1. Construction of the estimator

We follow the low-frequency literature [44, 65, 23] and consider a diffusion model on $[0,1]$ with boundary reflection (see Section 4.2.4 for a discussion of the model). Let $\|\cdot\|_\infty$ denote the supremum norm on space $B([0,1])$ of bounded measurable functions on $[0,1]$. Finally, for $i = 1, 2$ denote by

$$H^i = \Big\{ f \in L^2([0,1]) : \, f \text{ has } i \text{ weak derivatives with } f^{(j)} \in L^2([0,1]),\, j \leq i \Big\}$$

the L^2-Sobolev spaces on $[0,1]$ of order i. H^i is a Hilbert space with the norm

$$\|f\|_{H^i} = \sum_{j \leq i} \|f^{(i)}\|_{L^2}.$$

Assumption 4.1. *For given constants* $0 < d < D$ *suppose* $(\sigma, b) \in \Theta$*, where*

$$\Theta := \Theta(d, D) = \{(\sigma, b) \in H^1([0,1]) \times B([0,1]) : \|b\|_\infty \vee \|\sigma^2\|_{H^1} < D, \inf_{x \in [0,1]} \sigma^2(x) \geq d\}.$$

Let the process $(X_t, t \geq 0)$ be given by the following Skorokhod type stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + dK_t, \tag{4.1}$$
$$X_t \in [0,1] \text{ for every } t \geq 0,$$

where $(W_t, t \geq 0)$ is a standard Brownian motion and $(K_t, t \geq 0)$ is an adapted continuous process with finite variation, starting from 0, such that for every $t \geq 0$ we have $\int_0^t \mathbf{1}_{(0,1)}(X_s)dK_s = 0$. The Sobolev regularity of σ ensures that the SDE (4.1) has a unique strong solution, see [93, Theorem 4]. As shown in [44], X admits an invariant measure with Lebesgue density.

Assumption 4.2. *The initial condition x_0 is distributed with respect to the invariant measure μ on $[0,1]$, independently of the driving Brownian motion W.*

Under Assumption 4.2, the diffusion X is stationary and ergodic. We denote with $\mathbb{P}_{\sigma,b}$ the law of X on the canonical space Ω of continuous functions over the positive axis with values in $[0,1]$, equipped with the topology of the uniform convergence on compact sets and endowed with its $\sigma-$field $\mathcal{F}$. We denote with $\mathbb{E}_{\sigma,b}$ the corresponding expectation operator.

Definition 4.3. Denote by $\widehat{\mu}_N$ the empirical measure associated to the observed sample:

$$\widehat{\mu}_N = \frac{1}{2N}\delta_{\{X_0\}} + \frac{1}{N}\sum_{n=1}^{N-1}\delta_{\{X_{n\Delta}\}} + \frac{1}{2N}\delta_{\{X_{N\Delta}\}}.$$

The underweighting of the first and the last observations is asymptotically negligible, but has meaningful finite sample interpretation both in the low and high-frequency regimes (see remarks before the equation (4.4) and after Definition 4.12). By ergodicity, when the time horizon T of the observed sample grows to infinity, the empirical measure $\widehat{\mu}_N(dx)$ converges weakly to the stationary distribution $\mu(dx)$. When T is fixed, but the observation frequency increases, the empirical measure tends to the occupation measure μ_T of the path $(X_t, 0 \leq t \leq T)$ (see Definition 4.7).

Definition 4.4. For $J \in \mathbb{N}_+$, $j = 1, ..., J$, let $\mathbf{1}_j(x) = \mathbf{1}(\frac{j-1}{J} \leq x < \frac{j}{J})$ be the indicator function of the j^{th} sub-interval and

$$\psi_j(x) = \int_0^x \mathbf{1}_j(y)dy, \text{ for } j = 1, ..., J,$$
$$\psi_0(x) = 1.$$

Let $V_J = \text{span}\{\psi_j : j = 0, ..., J\}$ be the space of linear splines with knots at $\{0, \frac{1}{J}, \frac{2}{J}, ..., \frac{J-1}{J}, 1\}$ and $V_J^0 = \{v \in V_J : \int_0^1 v(x)\widehat{\mu}_N(dx) = 0\}$ be the subspace of functions $L^2(\widehat{\mu}_N)-$orthogonal to constants.

Consider the generalized symmetric eigenproblem:

Eigenproblem 4.5. Find $(\widehat{\gamma}, \widehat{u}) \in \mathbb{R} \times V_J$ with $\widehat{u} \neq 0$, such that

$$\widehat{l}(\widehat{u}, v) = \widehat{\gamma}\widehat{g}(\widehat{u}, v), \text{ for all } v \in V_J,$$

where $\widehat{g}, \widehat{l} : V_J \times V_J \to \mathbb{R}$ are symmetric, bilinear forms defined by:

$$\widehat{g}(u,v) = \int_0^1 u(x)v(x)\widehat{\mu}_N(dx),$$
$$\widehat{l}(u,v) = \frac{1}{2T}\sum_{n=0}^{N-1} \big(u(X_{(n+1)\Delta}) - u(X_{n\Delta})\big)\big(v(X_{(n+1)\Delta}) - v(X_{n\Delta})\big).$$

When the observed sample visits at least twice every interval $[\frac{j-1}{J}, \frac{j}{J})$, the form $\widehat{g}$ is positive definite on V_J, while $\widehat{l}$ is positive semi-definite on V_J and positive definite on V_J^0. In such a case, Eigenproblem 4.5 has $\dim(V_J) = J+1$ solutions $(\widehat{\gamma}_j, \widehat{u}_j)_{j=0,\dots,J}$, with non-negative eigenvalues $0 \le \widehat{\gamma}_0 \le \widehat{\gamma}_1 \le \dots \le \widehat{\gamma}_J$ and $\widehat{g}-$orthogonal eigenfunctions. It is easy to check that $\widehat{\gamma}_0 = 0$ is an eigenvalue which corresponds to the constant function. Since the eigenfunctions are $\widehat{g}-$orthogonal, it follows that $\widehat{u}_j \in V_J^0$ for $1 \le j \le J$. Consequently, $\widehat{\gamma}_1 > 0$.

Definition 4.6. Let

$$\widehat{\zeta}_1 = \frac{\log(1-\Delta\widehat{\gamma}_1)}{\Delta}\mathbf{1}(\Delta\widehat{\gamma}_1 < 1) \quad \text{and} \quad \widehat{u}_1(x) = \sum_{j=0}^{J} \widehat{u}_{1,j}\psi_j(x).$$

When $\widehat{u}_{1,j} \neq 0$ we define the spectral estimator by

$$\widehat{\sigma}^2_{S,j} = \frac{-2\widehat{\zeta}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\int_0^1 \psi_j'(x)\widehat{u}_{1,j}\widehat{\mu}_N(dx)},$$
$$\widehat{\sigma}^2_S(x) = \sum_{j=1}^{J} \widehat{\sigma}^2_{S,j}\mathbf{1}_j(x).$$

The condition $\mathbf{1}(\Delta\widehat{\gamma}_1 < 1)$ is a technical assumption which ensures that the estimator $\widehat{\zeta}_1$ is well defined. As explained in Section 4.2.1, $1-\Delta\widehat{\gamma}_1$ is the estimator of the largest nontrivial eigenvalue of the transition operator. When $\Delta\widehat{\gamma}_1 \ge 1$, the estimated transition operator is negative definite on V_J^0, thus the spectral approach will not provide a reliable output. Proposition 4.20 and Corollary (4.52) ensure that $\Delta\widehat{\gamma}_1 < 1$ with high probability, both in high and low-frequency regimes.

4.1.2. High-frequency convergence rate

The estimation of volatility at point x is possible only when the process spends enough time around x.

Definition 4.7. Set $T > 0$. Define the occupation density

$$\mu_T = \frac{L_T}{T\sigma^2}, \tag{4.2}$$

where L_T is the semimartingale local time of the path $(X_t : 0 \le t \le T)$.

For any bounded Borel measurable function f, the following occupation formula holds:

$$\frac{1}{T}\int_0^T f(X_s)ds = \int_0^1 f(x)\mu_T(x)dx. \tag{4.3}$$

In order to obtain the global rates of convergence, we must assume that the occupation density of the observed path is bounded from below. Therefore, for a given level v, we study the risk of the estimator conditioned to the event

$$\mathcal{L}_v = \left\{ \inf_{x\in[0,1]} \mu_T(x) \geq v \right\}.$$

Theorem 4.8. Grant Assumptions 4.1 and 4.2. Fix $T > 0$, $0 < a < b < 1$ and $v > 0$. Choose J such that $c^{-1}\Delta^{-1/3} \leq J \leq c\Delta^{-1/3}$ holds for some constant c. For every $\varepsilon > 0$ and $\Delta > 0$ sufficiently small, there exists an event $\mathcal{R}_{\varepsilon,\Delta}$, with $\mathbb{P}_{\sigma,b}(\mathcal{R}_{\varepsilon,\Delta}) \geq 1-\varepsilon$, and a positive constant $C = C(T, a, b, v, c, \varepsilon, d, D)$, such that

$$\sup_{(\sigma,b)\in\Theta(d,D)} \mathbb{E}_{\sigma,b}\left[\mathbf{1}_{\mathcal{R}_{\varepsilon,\Delta}\cap\mathcal{L}_v} \cdot \|\widehat{\sigma}^2_S - \sigma^2\|_{L^1([a,b])}\right] \leq C\Delta^{\frac{1}{3}}.$$

Hoffmann [49, Proposition 2] shows that the rate $\Delta^{1/3}$ is optimal in the minimax sense even in the class of diffusions with Lipschitz volatility. To prove Theorem 4.8, we compare $\widehat{\sigma}^2_S$ with the benchmark Florens-Zmirou estimator, see Section 4.2.2. While the consistency of the spectral estimator can be obtained using the well known path properties of diffusion processes, the proof of the exact convergence rate is rather demanding. As explained in Section 4.3.2, it is necessary to show the regularity properties of the estimated eigenfunction $\widehat{u}_1$, which requires rather sophisticated arguments from the perturbation theory of differential operators with non-smooth coefficients.

4.1.3. Low-frequency convergence rate

In the low-frequency regime, we need to threshold the estimator in order to ensure integrability and stability against large stochastic errors. As expected, $\widehat{\sigma}^2_S$ achieves the same mean L^2 rate as the original Gobet-Hoffmann and Reiß estimator. Furthermore, for $\sigma \in H^1$, this rate is minimax optimal, which can be obtained by the same proof as [44, Theorem 2.5].

Theorem 4.9. Grant Assumptions 4.1 and 4.2. Fix $\Delta > 0$ and $0 < a < b < 1$. Choose J such that $c^{-1}N^{1/5} \leq J \leq cN^{1/5}$ holds for some constant c. We have

$$\sup_{(\sigma,b)\in\Theta(d,D)} \mathbb{E}_{\sigma,b}\left[\|\widehat{\sigma}^2_S \wedge D - \sigma^2\|^2_{L^2([a,b])}\right]^{\frac{1}{2}} \leq CN^{-\frac{1}{5}},$$

where the constant C depends on Δ, a, b, c, d, D but not on the sample size N.

The general idea of the proof is the same as in Gobet et al. [44] or [23]. We use the mixing property of the process X to control the approximation error of the stationary measure μ by the empirical measure $\widehat{\mu}_N$, see Corollary 4.45. Then, as discussed in Section 4.2.1, we bound the estimation error of (κ_1, u_1) - the first nontrivial eigenpair of the transition operator P_Δ, obtaining

$$|\widehat{\kappa}_1 - \kappa_1| + \|\widehat{u}_1 - u_1\|_{H^1} = O_{\mathbb{P}}(N^{-1/5}).$$

Finally, we bound the plug-in error of the spectral estimator $\widehat{\sigma}_S$. A tenuous point is in that the estimator $\widehat{u}_1$ converges to the eigenfunction u_1 in the sense of mean H^1 norm only, hence we can not postulate a uniform positive lower bound on $\inf_{x\in[a,b]} \widehat{u}_1'(x)$. Following Chorowski and Trabs [23], we are able to overcome this difficulty by applying the threshold $\widehat{\sigma}^2_S \wedge D$.

4.2. Discussion

4.2.1. Connection to the GHR low-frequency estimator

In this section, we explain the relation between the defined estimator $\widehat{\sigma}_S$ above and the original spectral estimator introduced in [44, Section 3.2]. First, let us review the construction of the GHR estimator.

Definition 4.10. As in Gobet et al. [44, Eq. 3.8] for $u, v \in V_J$ let

$$\widehat{p}(u,v) = \frac{1}{2N}\sum_{n=0}^{N-1}\big(u(X_{n\Delta})v(X_{(n+1)\Delta}) + v(X_{n\Delta})u(X_{(n+1)\Delta})\big).$$

A crucial observation is that, due to the appropriate weighting of the empirical measure, $\widehat{p}$ becomes a linear combination of $\widehat{l}$ and $\widehat{g}$. Indeed, using the summation by parts formula, we obtain

$$\widehat{l} = \frac{1}{\Delta}(\widehat{g} - \widehat{p}). \tag{4.4}$$

Hence, for $(\widehat{\gamma}_i, \widehat{u}_i)$- any solution of the Eigenproblem 4.5, we have

$$\widehat{p}(\widehat{u}_i, v) = (1 - \Delta\widehat{\gamma}_i)\widehat{g}(\widehat{u}_i, v) \text{ for every } v \in V_J. \tag{4.5}$$

Denote

$$\widehat{\kappa}_i = (1 - \Delta\widehat{\gamma}_i). \tag{4.6}$$

We conclude that the eigenpair $(\widehat{\kappa}_1, \widehat{u}_1)$ is equal to the estimator of the eigenpair of the transition operator which is defined in [44, Eq 3.11]. Taking into account that functions (ψ_j) are not orthonormal, following [44, Eq. 3.12 and Eq. 3.7], we define the GHR estimator as:

Definition 4.11.

$$\widehat{\sigma}^2_{GHR}(x) = \frac{2\widehat{\zeta}_1\int_0^x \widehat{u}_1(y)\widehat{\mu}_N(dy)}{\widehat{u}_1'(x)\widehat{\mu}(x)},$$

where

$$\widehat{\mu} = \sum_{j=0}^{J}\widehat{\mu}_j\psi_j \text{ with } (\widehat{\mu}_j)_j = \Big(\Big[\int_0^1 \psi_i(y)\psi_j(y)dy\Big]_{i,j}\Big)^{-1}\Big(\int_0^1 \psi_i(x)\widehat{\mu}_N(dx)\Big)_i,$$

is an estimator of the stationary density.

Note that the estimator $\widehat{\sigma}^2_S$ can be seen as a local average of $\widehat{\sigma}^2_{GHR}$. Indeed, since $\mathbf{1}_j = \psi_j'$, integrating by parts gives us

$$\widehat{\sigma}^2_{S,j} = \frac{2\widehat{\zeta}_1\int_0^1 \psi_j'(x)\Big(\int_0^x \widehat{u}_1(y)\widehat{\mu}_N(dy)\Big)dx}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{u}_1'(x)\widehat{\mu}_N(dx)} = \frac{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\sigma}^2_{GHR}(x)\widehat{u}_1'(x)\widehat{\mu}(x)dx}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{u}_1'(x)\widehat{\mu}_N(dx)}. \tag{4.7}$$

Since we focus on volatility functions in H^1, the above averaging has no effect on the low-frequency convergence rate. On the other hand, there are multiple reasons why it is beneficial for optimality in the high-frequency regime. Firstly, since $\widehat{u}_1'$ is constant on every interval

$\left[\frac{j-1}{J}, \frac{j}{J}\right]$, after averaging we do not have to estimate the density of the occupation measure (which is not regular in the high-frequency setting), but the occupation measure of the intervals $\left[\frac{j-1}{J}, \frac{j}{J}\right]$. Furthermore, averaging reduces the variance of the estimator, which can be clearly seen in Figure 4.1. The intuitive explanation of this phenomenon is that while the original estimator $\widehat{\sigma}^2_{GHR}$ inherits the rough behaviour of the occupation density (via the inverse of the derivative of the eigenfunction u_1 which has the same smoothness as the design density) this irregularity is removed by multiplication with $\widehat{u}_1'\widehat{\mu}$.

4.2.2. Connection to the Florens-Zmirou estimator

The general idea of the proof of the high-frequency convergence rate is to compare estimator $\widehat{\sigma}_S$ with the minimax optimal (see [49, Proposition 2]) high-frequency estimator introduced in Florens-Zmirou [38]. In this section, we recall the definition of the Florens-Zmirou estimator and discuss its relation to $\widehat{\sigma}_S$.

Definition 4.12. Define the time-symmetric version of the well known Nadaraya-Watson type estimator of the squared volatility coefficient, introduced in Florens-Zmirou [38], by

$$\begin{aligned} \widehat{\sigma}^2_{FZ,j} &= \frac{\sum_{n=0}^{N-1}(\mathbf{1}_j(X_{n\Delta}) + \mathbf{1}_j(X_{(n+1)\Delta}))(X_{(n+1)\Delta} - X_{n\Delta})^2}{\Delta\sum_{n=0}^{N-1}(\mathbf{1}_j(X_{n\Delta}) + \mathbf{1}_j(X_{(n+1)\Delta}))}, \\ \widehat{\sigma}^2_{FZ}(x) &= \sum_{j=1}^{J} \widehat{\sigma}^2_{FZ,j}\mathbf{1}_j(x). \end{aligned}$$

Note that the underweighting of the first and last observation in the denominator of $\widehat{\sigma}^2_{FZ,j}$ appears naturally as an artifact of the time symmetry.

Remark 4.13. We call $\widehat{\sigma}^2_{FZ}$ a time-symmetrized version of the Florens-Zmirou estimator, since it is an average of the standard Florens-Zmirou estimators (c.f. [38, Eq. (1.1)]) constructed for the process $(X_t, 0 \le t \le T)$ and the time reversed process $Y_t = X_{T-t}$. Indeed, let

$$\widehat{\sigma}^2_j(X_0, X_\Delta, ..., X_{N\Delta}) = \frac{\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})(X_{(n+1)\Delta} - X_{n\Delta})^2}{\Delta(\frac{1}{2}\mathbf{1}_j(X_0) + \sum_{n=1}^{N-1}\mathbf{1}_j(X_{n\Delta}) + \frac{1}{2}\mathbf{1}_j(X_{N\Delta}))}. \tag{4.8}$$

Then

$$\widehat{\sigma}^2_{FZ,j} = \frac{\widehat{\sigma}^2_j(X_0, X_\Delta, ..., X_{N\Delta}) + \widehat{\sigma}^2_j(Y_0, Y_\Delta, ..., Y_{N\Delta})}{2}.$$

Since stationary scalar diffusions are reversible, under the Assumption 4.2, the process $(Y_t, 0 \le t \le T)$ is identical in law to $(X_t, 0 \le t \le T)$. Hence, the statistical properties of estimator $\widehat{\sigma}^2_{FZ}$ are the same as those of the classical Florens-Zmirou estimator.

Recall that $(\widehat{\gamma}_1, \widehat{u}_1)$ is an eigenpair of the Eigenproblem 4.5. From Definition 4.6 of the spectral estimator, it follows that

$$\widehat{\sigma}^2_{S,j} = \frac{-\widehat{\zeta}_1}{\widehat{\gamma}_1}\frac{2\widehat{l}(\widehat{u}_1, \psi_j)}{\widehat{u}_{1,j}\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}. \tag{4.9}$$

A similar representation formula can be established for the time symmetric Florens-Zmirou estimator $\widehat{\sigma}^2_{FZ}$.

Definition 4.14. Define a bilinear form $\widehat{f} : V_J \times V_J \to \mathbb{R}$ by

$$\widehat{f}(u,v) = \frac{1}{2}\int_0^1 u'(x)v'(x)\widehat{\sigma}^2_{FZ}(x)\widehat{\mu}_N(dx).$$

Consider vector $(v_j)_{j=1,...,J}$ such that $v_j \neq 0$ for every $j = 1,...,J$ and the associated function $v \in V_J^0$. We have

$$\widehat{\sigma}^2_{FZ,j} = \frac{2\widehat{f}(v,\psi_j)}{v_j \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)}. \tag{4.10}$$

As will be thoroughly explained in Section 4.3.2, when $\Delta \to 0$, the eigenvalue ratio $-\widehat{\zeta}_1/\widehat{\gamma}_1$ in (4.9) tends to 1. Consequently, the difference between estimators $\widehat{\sigma}^2_S$ and $\widehat{\sigma}^2_{FZ}$ is controlled by

$$\frac{2|\widehat{l}(\widehat{u}_1,\psi_j) - \widehat{f}(\widehat{u}_1,\psi_j)|}{\widehat{u}_{1,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)}. \tag{4.11}$$

The main observation is that in the high-frequency analysis, we do not have to control the estimation error of the derivative $\widehat{u}_1'$. Indeed, to bound (4.11), we need only to show a uniform lower bound for $\widehat{u}_{1,j}$ and an upper bound for the difference $|\widehat{l}(\widehat{u}_1,\psi_j) - \widehat{f}(\widehat{u}_1,\psi_j)|$. Unfortunately, $|\widehat{l}(v,\psi_j) - \widehat{f}(v,\psi_j)|$ is not small enough for any bounded function v. To achieve the required upper bound for the estimated eigenfunction, we need to first obtain some regularity properties of $\widehat{u}_1$, which is the most difficult part of the high-frequency analysis.

4.2.3. Numerical analysis

In this section, we present the numerical results for the volatility estimation across different observation time scales. We compare three estimation methods: the time symmetric Florens-Zmirou estimator $\widehat{\sigma}^2_{FZ}$ (see Definition 4.12), the spectral estimator $\widehat{\sigma}^2_{GHR}$ (see Definition 4.11, c.f. Gobet et al. [44, Section 3.2]) with approximation space V_J of linear splines with equidistant knots, and finally, the locally averaged spectral estimator $\widehat{\sigma}^2_S$. We apply an oracle choice of the projection level J, minimizing the risk. The source code used for presented simulations is available at [22].

We compare the locally averaged spectral estimator $\widehat{\sigma}^2_S$ with benchmark estimators $\widehat{\sigma}^2_{FZ}$ and $\widehat{\sigma}^2_{GHR}$ in both high and low-frequency regimes. Following Chorowski and Trabs [23, Section 5] we consider diffusion process X with mean reverting drift

$$b(x) = 0.2 - 0.4x, \tag{4.12}$$

quadratic squared volatility function

$$\sigma^2(x) = 0.4 - (x - 0.5)^2, \tag{4.13}$$

and two reflecting barriers at 0 and 1. This choice of diffusion coefficients is supposed to minimize the reflection effect alongside with some variability in the volatility function. Nevertheless, the depicted behaviour is typical for other diffusion processes. The sample paths were generated using the Euler-Maruyama scheme with time step size $\Delta/100 \wedge 0.001$ with reflection after each step. All simulated paths were conditioned to have an occupation time

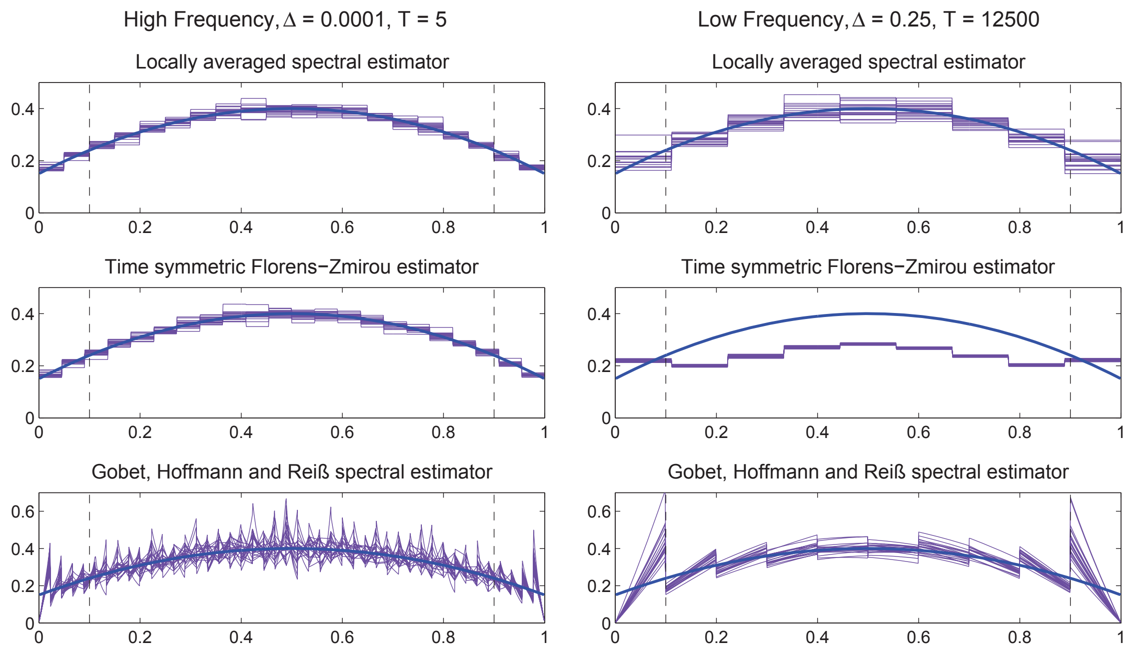

Figure 4.1.: Estimated squared volatility function for 20 independent trajectories. $\Delta = 0.0001$, $T = 5$, J is set to its oracle value.

density greater than $v = 0.2$. Table 4.1 presents the oracle mean $L^1([0.1, 0.9])$ estimation error of σ^2, obtained by a Monte Carlo simulation with 1000 iterations, in high ($T = 5, \Delta \to 0$) and low ($\Delta = 0.25, T \to \infty$) frequency regimes, respectively. The estimated squared volatility functions for 20 independent paths are depicted in Figure 4.1. Finally, in Figure 4.3 we present an example of the estimated occupation density, eigenfunction and its derivative.

In the case of high-frequency observations, $\widehat{\sigma}^2_S$ performs similarly to the benchmark estimator $\widehat{\sigma}^2_{FZ}$. Relative to $\|\sigma^2\|_{L^1([0.1,0.9])} \approx 0.28$, the error decreases from approximately 6% for $\Delta = 10^{-3}$ to 3% for $\Delta = 10^{-4}$. The estimation error of spectral estimator $\widehat{\sigma}^2_{GHR}$ is almost twice as large, although the quality of the estimation improves when Δ decreases. It is important to note that the oracle values of space parameter J for $\widehat{\sigma}^2_{GHR}$ are much bigger than those for other estimation methods. When Δ is small, the eigenfunctions inherit the regularity of the local time; the increase in dimension compensates for the projection error. Due to local averaging, this irregularity problem does not appear for $\widehat{\sigma}^2_S$, compare with Figure 4.1, where estimator $\widehat{\sigma}^2_{GHR}$ oscillates heavily. Furthermore, there is no visible boundary effect, suggesting that the error rate of the spectral estimator does not deteriorate outside the fixed interval $[0.1, 0.9]$.

In the low-frequency regime, $\widehat{\sigma}^2_S$ performs slightly better than the original spectral estimator $\widehat{\sigma}^2_{GHR}$. The boundary problem is visible, especially for $\widehat{\sigma}^2_{GHR}$. The relative error decreases from 12% for T=1000 to 5% for T=30 000. The Florens-Zmirou estimator $\widehat{\sigma}^2_{FZ}$ underestimates the volatility and commits a relative error of 30% . This is expected and due mostly to the boundary reflection, which, for low-frequency observations, is not negligible in the interior of the state space. As found by unreported simulations, in the case of low-frequency observations, the locally averaged spectral estimator $\widehat{\sigma}^2_S$ will outperform the Florens-Zmirou estimator in the case of a highly varying volatility function σ^2, even when the sampling frequency is big

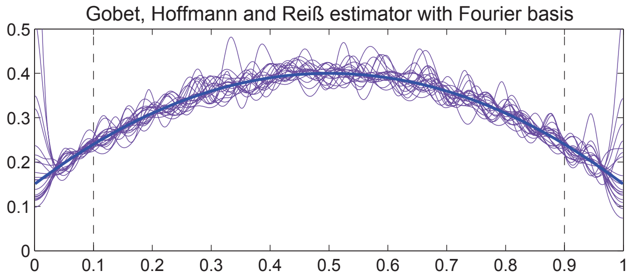

Figure 4.2.: Estimated squared volatility function using GHR estimator for 20 independent trajectories. $\Delta = 0.0001$, $T = 5$, J is set to its oracle value.

High-Frequency Regime: $T = 5$

	$\Delta = 0.001$	$\Delta = 0.00075$	$\Delta = 0.0005$	$\Delta = 0.00035$	$\Delta = 0.0002$	$\Delta = 0.0001$
$\widehat{\sigma^2_{GHR}}$	$0.0388_{(18)}$	$0.0353_{(23)}$	$0.0322_{(24)}$	$0.0292_{(32)}$	$0.0258_{(36)}$	$0.0220_{(49)}$
$\widehat{\sigma^2_S}$	$0.0195_{(9)}$	$0.0174_{(10)}$	$0.0149_{(10)}$	$0.0131_{(12)}$	$0.0108_{(13)}$	$0.0088_{(18)}$
$\widehat{\sigma^2_{FZ}}$	$0.0169_{(10)}$	$0.0153_{(11)}$	$0.0133_{(12)}$	$0.0119_{(12)}$	$0.0100_{(13)}$	$0.0080_{(20)}$

Low-Frequency Regime:$\Delta = 0.25$

	T=1k	T=3k	T=7k	T=10k	T=15k	T=20k
$\widehat{\sigma^2_{GHR}}$	$0.0386_{(5)}$	$0.0333_{(6)}$	$0.0256_{(11)}$	$0.0226_{(11)}$	$0.0198_{(11)}$	$0.0178_{(11)}$
$\widehat{\sigma^2_S}$	$0.0310_{(4)}$	$0.0245_{(6)}$	$0.0200_{(7)}$	$0.0182_{(8)}$	$0.0166_{(8)}$	$0.0155_{(9)}$
$\widehat{\sigma^2_{FZ}}$	$0.0821_{(5)}$	$0.0823_{(5)}$	$0.0823_{(5)}$	$0.0822_{(5)}$	$0.0823_{(5)}$	$0.0824_{(5)}$

Table 4.1.: Monte Carlo estimation errors in high and low-frequency regimes. The value of the parameter J is given in the subscript.

enough to ignore the reflection effect.

GHR estimator in the high-frequency regime

In this section, we want to analyze the performance of the GHR estimator in the high-frequency regime. Consider a diffusion process X with drift and volatility coefficients given by (4.12) and (4.13), respectively. Figure 4.2 depicts volatility functions estimated using the GHR estimator implemented with Fourier cosine basis. Corresponding oracle mean $L^1([0.1, 0.9])$ estimation errors, obtained by a Monte Carlo simulation with 1000 iterations, are presented in Table 4.2. All paths used in estimation were conditioned to have the chronological local time on $[0, 1]$ above the level $v = 0.2$.

The GHR estimator with Fourier basis exhibits the same characteristic features as when implemented with linear spline basis. The oracle values of the basis dimension are rather high, which is due to the roughness of the eigenfunction u_1. This is reflected in the shape of the estimated functions that fluctuate around the true squared volatility function. It is interesting that the empirical convergence rate (defined as the slope of the linear fit to the $\log(\Delta)\log(Error)$ plots) is close to the optimal (for H^1 volatility) rate 1/3. This might be because Fourier basis allows a better approximation of the smooth function σ^2 than piecewise constant approximation. The boundary effect is clearly visible, even that cosine functions satisfy Neumann boundary conditions.

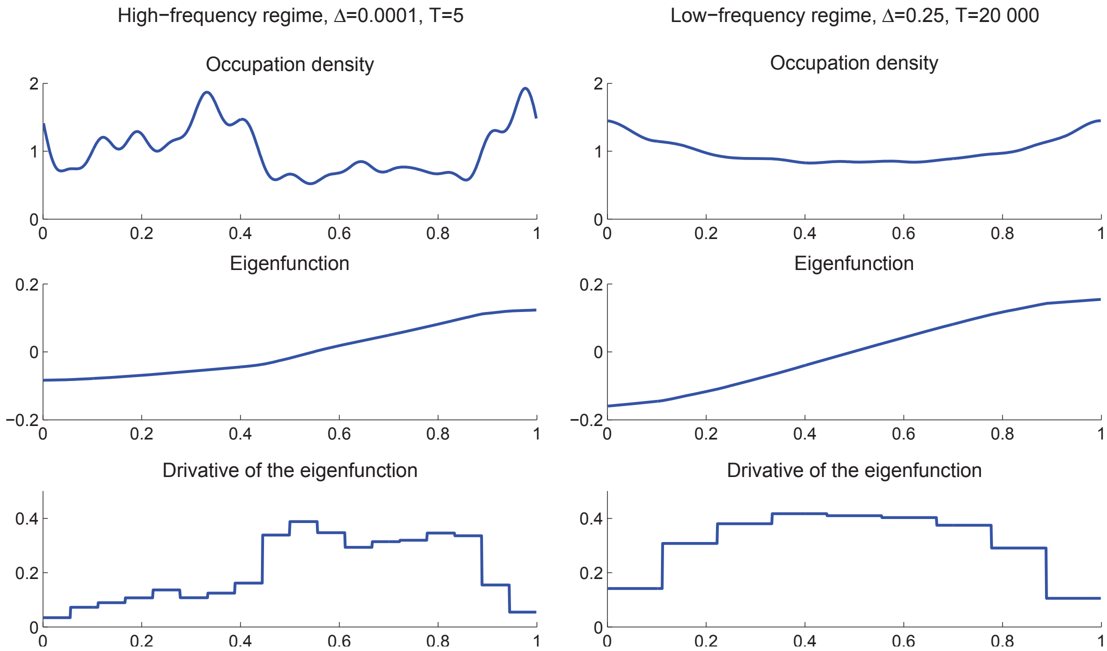

Figure 4.3.: An example of the estimated occupation density, eigenfunction and its derivative in high and low-frequency data. In the low frequency regime occupation density and the eigenfunction approximate the invariant density and the eigenfunction of the generator. In the high-frequency regime they are random objects that depend on the local time of the observed path.

High-Frequency Regime: $T = 5$

	$\Delta = 0.001$	$\Delta = 0.00075$	$\Delta = 0.0005$	$\Delta = 0.00035$	$\Delta = 0.0002$	$\Delta = 0.0001$	Empirical rate
Fourier basis	$0.0241_{(10)}$	$0.0221_{(12)}$	$0.0193_{(14)}$	$0.0169_{(16)}$	$0.0119_{(22)}$	$0.0114_{(28)}$	0.33

Table 4.2.: Monte Carlo estimation errors of the GHR estimator in high-frequency regime with Fourier cosine basis. The value of the parameter J is given in the subscript.

4.2.4. Extensions and limitations

Stationarity of process X. In the high-frequency analysis the stationarity assumption ensures that process X is time reversible. General initial distributions could be considered, but in order to preserve the performance of the estimation for the time reversed process, the coefficients of the backward process must belong to the nonparametric family Θ.

Due to the spectral gap of the generator, process X is geometrically ergodic. In particular, as $t \to \infty$, the one dimensional distributions of X_t converge exponentially fast to the invariant measure μ. It follows that, in the low-frequency regime, the assumption of stationarity can be made without loss of generality for asymptotic results.

Estimation at the boundaries. In the high-frequency regime, we prove the error bound in the interior of the state space. Restriction to the interval (a, b) allows us to obtain uniform lower bounds on the derivative of eigenfunction $\widehat{u}_1$, which, due to boundary conditions, are not valid in the entire state space. This restriction could be omitted by obtaining uniform bounds on the ratio of derivatives $\widehat{u}_{1,j\pm1}/\widehat{u}_{1,j}$. Unfortunately, since our proof relies on a posteriori error bounds on solutions for perturbed eigenvalue problems, we do not have the sufficient tools to control the pointwise relative error of the eigenfunctions. Nevertheless, the numerical results suggest that the spectral estimation procedure also behaves well at the boundaries of the state space.

In the low-frequency regime, the spectral estimator is unstable at the boundary due to Neumann boundary conditions for the eigenfunctions of the infinitesimal generator. Refer to [44, Section 3.3.8] for a discussion of the boundary problem.

Boundary reflection. Following previous works on the spectral estimation in the low frequency setting, e.g. [44, 65, 23], we consider an Itô diffusion model on the state space $[0,1]$ with instantaneous reflection at the boundaries. The assumption of a compact state space makes the construction of the estimator easier and facilitates error analysis in the low-frequency setting, c.f. Reiß [72]. We point out, here, that the reflection assumption is not restrictive in the high-frequency setting. Indeed, consider diffusion X defined on the entire real line with drift b and volatility σ. Let

$$A(t) = \int_0^t \mathbf{1}_{[0,1]}(X_s)ds$$

be the occupation time of interval $[0,1]$. Assume that $\lim_{t\to\infty} A(t) = \infty$ and define the right-continuous inverse

$$C(t) = \inf\{s > 0 | A(s) > t\}.$$

Process $Y_t = X_{C(t)}$ follows the law of a reflected diffusion on $[0,1]$ with drift b and volatility σ. Assume now the given observations $X_0, X_\Delta, ..., X_{N\Delta}$. The sub-sequence of the values that lie in $[0,1]$ forms a chain of observation of Y. The sampling frequency is random (and depends on the path), but when Δ shrinks, it becomes close to equidistant. The difficulty in handling irregularities at the boundaries is similar to these found when considering the reflection effect. Unfortunately, while this reduction can be used under the assumption that Δ is small, it cannot be applied in the low frequency setting, hence it is not practical in the context of scale invariant estimation.

Linear spline basis. The use of the linear spline basis is very convenient, as functions ψ_j appear naturally after applying integration by parts to the locally averaged GHR estimator, see (4.7). Nevertheless, simulation results presented in Section 4.2.3 suggest that the spectral estimation method performs as well with other bases. The Fourier cosines basis on $[0,1]$ is especially efficient, consisting of the eigenfunctions of the reflected Brownian motion process.

Adaptivity. An important decision in the spectral estimation is the choice of the basis dimension J. The general problem is twofold: dimension J should adapt to the smoothness of the coefficients and simultaneously to the observation frequency. In [23] the authors applied Lepski's method to construct a data-driven version of the GHR estimator that adapts to the smoothness of the volatility. In the case of the low frequency data, the same selection rule can be applied for the universal estimator $\widehat{\sigma}_S$. The precise construction of a method that will adapt to the observation frequency remains open.

The numerical study shows that the proposed estimator $\widehat{\sigma}_S$ smoothly interpolates between the high and low-frequency estimators. The optimal convergence rates in both frequency regimes leave out the question of the paradigm to use when one has to consider data. The different convergence rates in high and low frequency regimes raise the question of bivariate asymptotics with respect to both Δ and T. Nevertheless, because of the structural differences of the high and low-frequency data, we believe that such an analysis would be particularly challenging.

4.3. High-frequency analysis

We will write $f \lesssim g$ (resp. $g \gtrsim f$) when $f \leq C \cdot g$ for some universal constant $C > 0$. $f \sim g$ is equivalent to $f \lesssim g$ and $g \lesssim f$.

The proof of Theorem 4.8 is presented in Section 4.3.5 and is accomplished in several steps. In Section 4.3.3 we prove the convergence rate of the time-symmetric Florens-Zmirou estimator. Section 4.3.4 is devoted to the proof of Proposition 4.20 - the uniform bounds on the estimated eigenpair $(\widehat{\gamma}_1, \widehat{u}_1)$. In Section 4.3.6 we prove some technical results on the crossing intensity of the diffusion processes.

4.3.1. Preliminaries

From now on we take the Assumptions 4.1 and 4.2 as granted. Fix $0 < a < b < 1$ and the level $v > 0$. For simplicity, set $T = 1$. Let $J \sim \Delta^{-1/3}$.

Sobolev regularity of the volatility implies that it is $1/2-$Hölder continuous. Indeed, by the Cauchy-Schwarz inequality it holds

$$\sup_{x,y\in[0,1]} \frac{|\sigma(x) - \sigma(y)|}{|x-y|^{1/2}} = \sup_{x,y\in[0,1]} \frac{|\int_x^y \sigma'(z)dz|}{|x-y|^{1/2}} \leq \|\sigma\|_{H^1}. \tag{4.14}$$

Recall Definition 4.7 of the occupation density μ_T. Formula (4.2), together with (4.14), imply that μ_T inherits the regularity properties of the local time. In particular

Theorem 4.15. The function μ_1 is almost surely Hölder continuous of order α for every $\alpha < 1/2$. Moreover, for every $p \geq 1$, we have

$$\sup_{(\sigma,b)\in\Theta} \mathbb{E}_{\sigma,b}\Big[\sup_{x\in[0,1]} \mu_1^p(x)\Big] \quad < \quad \infty. \tag{4.15}$$

$$\sup_{(\sigma,b)\in\Theta} \mathbb{E}_{\sigma,b}\big[|\mu_1(x) - \mu_1(y)|^{2p}\big] \quad \leq \quad C_p|x-y|^p. \tag{4.16}$$

Proof. Since σ is uniformly bounded and $1/2-$Hölder continuous, the claim of the theorem can be deduced from the well known properties of the family of the local times $(L_t, t \geq 0)$ of the semimartingale X, see the proof of [73, Chapter VI, Theorem 1.7] and the subsequent remark. □

Definition 4.16. Denote by ω the modulus of continuity of the path $(X_t, 0 \leq t \leq 1)$, i.e.

$$\omega(\delta) = \sup_{0\leq s,t\leq 1,\, |t-s|\leq\delta} |X_t - X_s|.$$

Because of the ellipticity assumption $\sigma > 0$ the path $(X_t, 0 \leq t \leq 1)$ shares the properties of Brownian paths. In particular, we can apply the Brownian upper bounds (see Fischer and

Nappo [36]) on the moments of ω:

Theorem 4.17. For every $p \geq 1$ there exists a constant $C_p > 0$ such that

$$\sup_{(\sigma,b)\in\Theta} \mathbb{E}_{\sigma,b}[\omega^p(\Delta)] \leq C_p \Delta^{p/2} ln^p \left(\Delta^{-1}\right). \tag{4.17}$$

For the proof of Theorem 4.17 refer to Chapter (2), Theorem (2.6). Using (4.17) we can show that on $\mathcal{L}_v$ the occupation measure $\widehat{\mu}_N$ is spread uniformly on $[0,1]$ with high probability.

Lemma 4.18. *Let*

$$\mathcal{R}_1 = \mathcal{L}_v \cap \{\omega(\Delta)\|\mu_1\|_\infty \leq \Delta^{5/11} v\}. \tag{4.18}$$

For Δ sufficiently small we have

$$\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}_1) \lesssim \Delta^{2/3}.$$

Furthermore, on the event $\mathcal{R}_1$, for every $1 \leq j \leq J$, we have

$$v \lesssim J \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) \lesssim \|\mu_1\|_\infty.$$

The proof of Lemma 4.18 is postponed to Section 4.3.3.

As mentioned in Section 4.1.2 we want to compare the spectral estimator $\widehat{\sigma}^2_S$ with the benchmark high-frequency estimator $\widehat{\sigma}^2_{FZ}$. Before that, we have to prove a uniform upper bound on the mean L^2 error of the time symmetric Florens-Zmirou estimator. The result below is a generalization of [49, Proposition 2], where the same rate was obtained under the assumptions of smooth drift and Lipschitz volatility. As proved in [49, Proposition 2] the rate $\Delta^{1/3}$ is optimal in the minimax sense even on the class of diffusions with Lipschitz volatility.

Theorem 4.19. Grant Assumptions 4.1 and 4.2. Fix $T > 0$ and choose $J \sim \Delta^{-\frac{1}{3}}$. We have

$$\sup_{(\sigma,b)\in\Theta(d,D)} \mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot \|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[1/J,1-1/J]}\big]^{\frac{1}{2}} \lesssim \Delta^{1/3}. \tag{4.19}$$

Because of the reflection, the rate deteriorates at the boundary. For $x \in [0,1/J] \cup [1-1/J,1]$

$$\sup_{(\sigma,b)\in\Theta(d,D)} \mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot |\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)|^2\big]^{\frac{1}{2}} \lesssim \Delta^{1/33}. \tag{4.20}$$

The proof of Theorem 4.19 is postponed to Section 4.3.3. The main idea is the decomposition of the error into a martingale and deterministic approximation parts as in [49, Proposition 2]. As expected, under the high-frequency assumption, the reflection has an effect only at the boundary. Inequalities (4.19) and (4.20) imply

$$\sup_{(\sigma,b)\in\Theta(d,D)} \mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot \|\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)\|_{L^1[0,1]}\big] \lesssim \Delta^{1/3}.$$

4.3.2. Outline of the proof of the high-frequency convergence rate

Since by Theorem 4.19 the estimator $\widehat{\sigma}^2_{FZ}$ attains the optimal rate $\Delta^{1/3}$, to prove Theorem 4.8 it is enough to upper bound the mean $L^1[0,1]$ error between $\widehat{\sigma}^2_{FZ}$ and $\widehat{\sigma}^2_S$. Using representations (4.9) and (4.10) $\widehat{\sigma}^2_{FZ} - \widehat{\sigma}^2_S$ can be reduced to the difference of the forms $\widehat{f}$ and $\widehat{l}$

(c.f. Lemma 4.35). First, we need however to list the properties of the eigenpair $(\widehat{\zeta}_1, \widehat{u}_1)$. The proof of the next Proposition is postponed to Section 4.3.4.

Proposition 4.20. *Let $0 < a < b < 1$ be fixed. For every $\varepsilon > 0$ there exists an event $\mathcal{R}_2 = \mathcal{R}_2(\varepsilon, \Delta)$, with $\mathbb{P}_{\sigma,b}(\mathcal{L}_v \backslash \mathcal{R}_2) \leq \varepsilon$, and a constant $C = C(\varepsilon)$, such that, for Δ sufficiently small we have*

$$\mathbf{1}_{\mathcal{R}_2} \cdot |\widehat{\gamma}_1| \lesssim C. \tag{4.21}$$

Furthermore, the eigenfunction $\widehat{u}_1$ can be chosen such that on $\mathcal{R}_2$

$$\sum_{j=1}^{J} \widehat{u}_{1,j}^2 = J \text{ and } \widehat{u}_{1,j} \sim 1, \text{ and } \sum_{j=1}^{J} \widehat{u}_{1,j}^2 \mathbf{1}(\widehat{u}_{1,j} < 0) \lesssim 1$$

hold for any $j = \lfloor aJ \rfloor - 1, ..., \lceil bJ \rceil + 1$.

Remark 4.21. The normalization $\sum_{j=1}^{J} \widehat{u}_{1,j}^2 = J$ is natural, as it is equivalent to $\|\widehat{u}_1'\|_{L^2} = 1$. In short, Proposition 4.20 states the existence of uniform bounds on $\widehat{u}_1' \mid_{[a,b]}$. Because of the Neumann boundary conditions on the generator, the separation from the boundary is necessary for the existence of the lower bound.

Remark 4.22. From the general inequality

$$|1 + \log(1-x)/x| \leq x, \quad 0 < x < 1/2,$$

together with the uniform bound (4.21) on the eigenvalue $\widehat{\gamma}_1$, we deduce that, on the high probability event $\mathcal{R}_2$, $|1 + \widehat{\zeta}_1/\widehat{\gamma}_1| \lesssim \Delta$ holds. Consequently, the eigenvalue ratio $-\widehat{\zeta}_1/\widehat{\gamma}_1$ in (4.9) is of no importance in the high-frequency analysis.

Definition 4.23. Define

$$\begin{aligned} \widetilde{\sigma}_{S,j}^2 &= \frac{2\widehat{l}(\widehat{u}_1, \psi_j)}{\widehat{u}_{1,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)}, \qquad (4.22) \\ \widetilde{\sigma}_S^2(x) &= \sum_{j=1}^{J} \widetilde{\sigma}_{S,j}^2 \mathbf{1}_j(x). \end{aligned}$$

For simplicity, we will refer from now on to $\widetilde{\sigma}_S$ as to the spectral estimator. Comparing the representations (4.22) and (4.10) we obtain

$$|\widetilde{\sigma}_{S,j}^2 - \widehat{\sigma}_{FZ,j}^2| = \frac{2|\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)|}{\widehat{u}_{1,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)}.$$

Since by Proposition 4.20 the derivative $\widehat{u}_{1,j}$ has a uniform lower bound, Lemma 4.18 implies that to show the convergence rate $\Delta^{1/3}$ we have to prove that

$$|\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)| = O_p(\Delta^{2/3}).$$

As argued in Proposition 4.36, for any function $v \in V_J$ with bounded derivative, it holds

$$|\widehat{l}(v, \psi_j) - \widehat{f}(v, \psi_j)| = O_p(\Delta^{1/2}),$$

which leads to a suboptimal rate $\Delta^{1/6}$. In order to achieve the optimal rate $\Delta^{1/3}$ we need to use the regularity of the first nontrivial eigenfunction $\widehat{u}_1$. By the means of the Perron-Frobenius theory, in Proposition 4.37, we prove that for some high probability event $\mathcal{R}_3$

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3}\cdot\Big|\frac{\widehat{u}_1'(\frac{j}{J}\pm\frac{1}{J})-\widehat{u}_1'(\frac{j}{J})}{J^{-1/2}}\Big|^2\Big]^{\frac{1}{2}}\lesssim 1$$

holds, which can be interpreted as the almost $1/2-$Hölder regularity of $\widehat{u}_1'$ (see Remark 4.38). This regularity of the eigenfunction allows us to reduce the estimation error to an approximation problem of the occupation time, see decomposition (4.59) and Lemma 4.39.

4.3.3. Proof of Theorem 4.19

We begin with the proof of Lemma 4.18.

Proof of Lemma 4.18. Note first that, on the event $\mathcal{L}_v$, we have

$$v\leq J\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(dx)\leq\|\mu_1\|_\infty.$$

Using the occupation formula (4.3), we obtain that

$$\begin{aligned}\Big|\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(x)dx\Big| &\leq\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}|\mathbf{1}_j(X_{n\Delta})-\mathbf{1}_j(X_s)|ds\leq\\ &\leq\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}\Big(\mathbf{1}(|X_s-\tfrac{j-1}{J}|<\omega(\Delta))ds+\mathbf{1}(|X_s-\tfrac{j}{J}|<\omega(\Delta))\Big)ds\\ &=\int_{\frac{j-1}{J}-\omega(\Delta)}^{\frac{j-1}{J}+\omega(\Delta)}\mu_1(x)dx+\int_{\frac{j}{J}-\omega(\Delta)}^{\frac{j}{J}+\omega(\Delta)}\mu_1(x)dx\leq 4\omega(\Delta)\|\mu_1\|_\infty.\end{aligned}$$

Hence, and since $J\sim\Delta^{-1/3}$, on the event $\mathcal{R}_1$

$$\Delta^{\frac{1}{3}}v\lesssim\Delta^{\frac{1}{3}}v-4\omega(\Delta)\|\mu_1\|_\infty\lesssim\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\lesssim(\Delta^{\frac{1}{3}}+4\omega(\Delta))\|\mu_1\|_\infty\lesssim\Delta^{\frac{1}{3}}\|\mu_1\|_\infty,$$

holds for any $\Delta<1$. Finally, to prove that $\mathcal{R}_1$ is a high probability event, note that for any $p\geq 1$, Theorem 4.17 together with the inequality (4.15) imply

$$\begin{aligned}\mathbb{P}_{\sigma,b}(\mathcal{L}_v\setminus\mathcal{R}_1) &\lesssim \Delta^{-5p/11}\mathbb{E}_{\sigma,b}[\omega(\Delta)^p\|\mu_1\|_\infty^p]\\ &\lesssim \Delta^{-5p/11}\mathbb{E}_{\sigma,b}[\omega(\Delta)^{2p}]^{1/2}\mathbb{E}_{\sigma,b}[\|\mu_1\|_\infty^{2p}]^{1/2}\\ &\lesssim \Delta^{-5p/11}\Delta^{p/2}\ln^{p/2}(\Delta^{-1}).\end{aligned}$$

We obtain the claim by choosing $p\geq 15$. □

Now, we are ready to prove Theorem 4.19. The main ideas are as in [49, Proposition 2]. The novelty consists on the direct treatment of the drift term and the analysis of the boundary behaviour, which is an artifact of the reflection.

Proof of Theorem 4.19. Set $\mathcal{R} = \mathcal{R}_1$. Recall the definition (4.8) and the discussion thereafter. It follows, that it is sufficient to prove the claim for $\widehat{\sigma}_j^2(X_0, X_\Delta, ..., X_{N\Delta})$.

Since

$$\Delta\Big(\frac{1}{2}\mathbf{1}_j(X_0) + \sum_{n=1}^{N-1} \mathbf{1}_j(X_{n\Delta}) + \frac{1}{2}\mathbf{1}_j(X_{N\Delta})\Big) = \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx),$$

by Lemma 4.18, on the event $\mathcal{R}_1$, the denominator of $\widehat{\sigma}_j^2(X_0, X_\Delta, ..., X_{N\Delta})$ has a uniform lower bound of order $\Delta^{1/3}$. Hence, in order to prove (4.19), we have to show that, for any $j = 2, ..., J-1$ and $x \in [\frac{j-1}{J}, \frac{j}{J}]$, we have

$$\begin{aligned}\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \big|\sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta})\big((X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)\big)\big|^2\Big]^{1/2} \\ \lesssim \Delta^{1/2}\Big(\int_{\frac{(j-3)}{J}\vee 0}^{\frac{(j+2)}{J}\wedge 1} [(\sigma^2)'(y)]^2 dy\Big)^{1/2} + \Delta^{2/3}.\end{aligned} \tag{4.23}$$

Indeed, (4.23) implies

$$\begin{aligned}\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[1/J,1-1/J]}\Big] &\lesssim \sum_{j=2}^{J-1} \frac{1}{J\Delta^{\frac{2}{3}}}\Big(\Delta\int_{\frac{(j-3)}{J}\vee 0}^{\frac{(j+2)}{J}\wedge 1} [(\sigma^2)'(y)]^2 dy + \Delta^{\frac{4}{3}}\Big) \\ &\lesssim \Delta^{2/3}(\|\sigma^2\|^2_{H^1} + 1).\end{aligned}$$

Step 1. Error bound in the interior. Fix $2 \le j \le J-1$ and $x \in [\frac{j-1}{J}, \frac{j}{J}]$. Note that on the event $\mathcal{R}_1$ the condition $\mathbf{1}_j(X_{n\Delta}) = 1$ implies that no reflection occurs for $t \in [n\Delta, (n+1)\Delta]$. Using Itô formula we can decompose

$$\sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta})\big((X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)\big) := A_1 + A_2 + A_3 + A_4,$$

where

$$\begin{aligned}A_1 &= \sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta})\Big[\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma(X_s)dW_s\Big)^2 - \int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds\Big],\\ A_2 &= \sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta}) \int_{n\Delta}^{(n+1)\Delta} (\sigma^2(X_s) - \sigma^2(x))ds,\\ A_3 &= \sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta})\Big(\int_{n\Delta}^{(n+1)\Delta} b(X_s)ds\Big)^2,\\ A_4 &= -2\sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta}) \int_{n\Delta}^{(n+1)\Delta} \sigma(X_s)dW_s \int_{n\Delta}^{(n+1)\Delta} b(X_s)ds.\end{aligned}$$

We will bound the second moment of each of the terms $A_1, ..., A_4$. First, note that arguing as in the proof of Lemma 4.18, we obtain

$$\frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}_j(X_{n\Delta}) \le (\Delta^{\frac{1}{3}} + 4\omega(\Delta))\|\mu_1\|_\infty. \tag{4.24}$$

Consequently, from the Cauchy-Schwarz inequality, together with Theorem 4.17 and the inequality (4.15) it follows that

$$\mathbb{E}_{\sigma,b}\Big[\Big(\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\Big)^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{3}}. \tag{4.25}$$

Denote by $\mathcal{F}_n$ the $\sigma-$field generated by $\{X_s : 0 \le s \le n\Delta\}$. Let

$$\eta_n = \Big(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)dW_s\Big)^2 - \int_{n\Delta}^{(n+1)\Delta}\sigma^2(X_s)ds.$$

Since $(\eta_n)_n$ are $(\mathcal{F}_n)-$martingale increments, they are conditionally uncorrelated. Using the Burkholder-Davis-Gundy inequality we obtain that $\mathbb{E}_{\sigma,b}[\eta_n^2|\mathcal{F}_n] \lesssim \Delta^2$. Consequently,

$$\mathbb{E}_{\sigma,b}[A_1^2]^{\frac{1}{2}} = \Big(\sum_{n=0}^{N-1}\mathbb{E}_{\sigma,b}[\mathbf{1}_j(X_{n\Delta})\eta_n^2]\Big)^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{2}}\mathbb{E}_{\sigma,b}\Big[\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{2}{3}},$$

where we used (4.25) to obtain the last inequality. On the event $\mathcal{R}_1$, when $\mathbf{1}_j(X_{n\Delta}) = 1$, we have

$$|X_s - x| \le |X_s - X_{n\Delta}| + |X_{n\Delta} - x| \le \omega(\Delta) + J^{-1} \le \Delta^{5/11} + J^{-1} \le 2J^{-1},$$

for $n\Delta \le s \le (n+1)\Delta$, $x \in [\frac{j-1}{J}, \frac{j}{J}]$ and Δ such that $\Delta^{5/11-1/3} \le c^{-1}$ (recall that $\Delta^{1/3} \le cJ^{-1}$). Hence,

$$\begin{aligned}
A_2 &\le \sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\int_{n\Delta}^{(n+1)\Delta}\Big|\int_x^{X_s}(\sigma^2)'(y)dy\Big|ds \le \sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\Delta\Big|\int_{\frac{(j-3)}{J}\vee 0}^{\frac{(j+2)}{J}\wedge 1}(\sigma^2)'(y)dy\Big| \\
&\lesssim \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\Delta^{1/6}\Big(\int_{\frac{(j-3)}{J}\vee 0}^{\frac{(j+2)}{J}\wedge 1}[(\sigma^2)'(y)]^2dy\Big)^{1/2},
\end{aligned}$$

where we used the Cauchy-Schwarz inequality. The above implies that

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_1}\cdot A_2^2]^{\frac{1}{2}} \le \Delta^{1/2}\Big(\int_{\frac{(j-3)}{J}\vee 0}^{\frac{(j+2)}{J}\wedge 1}[(\sigma^2)'(y)]^2dy\Big)^{1/2}.$$

The drift function b is uniformly bounded, hence $|A_3| \lesssim \Delta$. Denote

$$Y_t = \int_0^t \sigma(X_s)dW_s \text{ and } \omega_Y(\Delta) = \sup_{\substack{0\le s,t\le 1\\ |t-s|\le\Delta}}|Y_t - Y_s|.$$

The uniform bound on b, together with $|Y_{(n+1)\Delta} - Y_{n\Delta}| \le \omega_Y(\Delta)$, and the inequality (4.24) implies

$$\begin{aligned}
\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_1}\cdot A_4^2]^{\frac{1}{2}} &\lesssim \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1}\cdot\Big(\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_j(X_{n\Delta})\omega_Y(\Delta)\Big)^2\Big]^{\frac{1}{2}} \\
&\lesssim \mathbb{E}_{\sigma,b}\Big[\big(\Delta^{\frac{1}{3}}\|\mu_1\|_\infty\omega_Y(\Delta)\big)^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{2}{3}},
\end{aligned}$$

where we used uniform bounds on the moments of modulus of continuity of semimartingales with bounded coefficients (see Theorem 4.17).

Step 2. Error bound at the boundaries. Set $j = 1$ (the case $j = J$ follows analogously) and $x \in [0, 1/J]$. On $\mathcal{R}_1$, whenever $X_{n\Delta} \geq \Delta^{5/11}$, no reflection occurs for $t \in [n\Delta, (n+1)\Delta]$. Denote

$$\mathbf{1}_1(x) = \mathbf{1}(x < \Delta^{5/11}) + \mathbf{1}(\Delta^{5/11} \leq x < J^{-1}) := \mathbf{1}_{1,0}(x) + \mathbf{1}_{1,1}(x).$$

We decompose

$$\sum_{n=0}^{N-1} \mathbf{1}_1(X_{n\Delta})\big((X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)\big) := E_1 + E_2,$$

with

$$\begin{aligned} E_1 &= \sum_{n=0}^{N-1} \mathbf{1}_{1,0}(X_{n\Delta})\big((X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)\big), \\ E_2 &= \sum_{n=0}^{N-1} \mathbf{1}_{1,1}(X_{n\Delta})\big((X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)\big). \end{aligned}$$

On $\mathcal{R}_1$ holds $|(X_{(n+1)\Delta} - X_{n\Delta})^2 - \Delta\sigma^2(x)| \lesssim \Delta^{10/11}$. Hence, arguing as in the proof of Lemma 4.18, we obtain that

$$\begin{aligned} \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot E_1^2\Big]^{\frac{1}{2}} &\lesssim \Delta^{-\frac{1}{11}} \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \Big(\frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}_{1,0}(X_{n\Delta})\Big)^2\Big]^{\frac{1}{2}} \\ &\lesssim \Delta^{-\frac{1}{11}} \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \Big(\int_0^1 \mathbf{1}_{1,0}(x)\mu_1(x)dx + 4\omega(\Delta)\|\mu_1\|_\infty\Big)^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{4}{11}}. \end{aligned}$$

To bound the second moment of E_2, note that when $\mathbf{1}_{1,1}(X_{n\Delta}) = 1$ no reflection occurs for $t \in [n\Delta, (n+1)\Delta]$. Consequently, we can proceed as in Step 1, obtaining

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot E_2^2\Big]^{\frac{1}{2}} \lesssim \Delta^{1/2}\Big(\int_0^{\frac{3\wedge J}{J}} [(\sigma^2)'(y)]^2 dy\Big)^{\frac{1}{2}} + \Delta^{2/3} \lesssim \Delta^{1/2}\|\sigma^2\|_{H^1}.$$

We conclude that

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot |\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{-1/3}\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot (E_1 + E_2)^2\Big]^{1/2} \lesssim \Delta^{1/33}.$$

$\square$

Corollary 4.24. *For every $\varepsilon > 0$ and Δ small enough there exists an event $\mathcal{R} = \mathcal{R}(\varepsilon, \Delta) \subseteq \mathcal{R}_1$, with $\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}) \leq \varepsilon$, such that on $\mathcal{R}$*

$$\forall x \in [0,1] \qquad (1 - 2^{-1/2})d \leq \widehat{\sigma}^2_{FZ}(x) \leq (1 + 2^{-1/2})D. \tag{4.26}$$

Proof. Since

$$\|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[0,1]} = \|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[1/J,1-1/J]} + \int_{[0,1/J]\cup[(J-1)/J,1]} |\sigma^2_{FZ}(x) - \sigma^2(x)|^2 dx,$$

using Theorem 4.19 we obtain that

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot \|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[0,1]}\big] \lesssim \Delta^{2/3} + J^{-1}\Delta^{2/33} \lesssim \Delta^{1/3+2/33}.$$

Set $\varepsilon > 0$. Let

$$\mathcal{R} = \mathcal{R}_1 \cap \big\{\|\widehat{\sigma}^2_{FZ} - \sigma^2\|^2_{L^2[0,1]} \leq (2J)^{-1} \inf_{x\in[0,1]} \sigma^4(x)\big\}.$$

From Markov's inequality, together with the lower bound on the probability of the event $\mathcal{R}_1$, it follows that

$$\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}) \lesssim \Delta^{2/33}.$$

Hence, for Δ sufficiently small, we have $\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}) \geq 1 - \varepsilon$. Fix $j = 1, ..., J$. On the event $\mathcal{R}$ we have

$$\begin{aligned}\Big|\widehat{\sigma}^2_{FZ,j} - J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x)dx\Big| &\leq J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{\sigma}^2_{FZ,j} - \sigma^2(x)|dx \leq J^{1/2}\left(\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\sigma^2(x) - \widehat{\sigma}^2_{FZ,j}|^2 dx\right)^{\frac{1}{2}}\\ &\leq J^{1/2}\|\widehat{\sigma}^2_{FZ} - \sigma^2\|_{L^2[0,1]} \leq 2^{-1/2} \inf_{x\in[0,1]} \sigma^2(x)\\ &\leq 2^{-1/2}J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x)dx.\end{aligned}$$

Since Assumption 4.1 implies that $d \leq J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x)dx \leq D$, we conclude that the claim follows. $\square$

4.3.4. Properties of the eigenpair $(\widehat{\gamma}_1, \widehat{u}_1)$

In this section we want to prove Proposition 4.20. Because of the tridiagonal structure of the form $\widehat{l}$, the direct analysis of the eigenfunction $\widehat{u}_1$ is difficult. Instead, we consider the generalized eigenvalue problem for forms $\widehat{f}$ (recall Definition 4.14) and $\widehat{g}$:

Eigenproblem 4.25. Find $(\widehat{\lambda}, \widehat{w}) \in \mathbb{R} \times V_J^0$, with $\widehat{w} \neq 0$, such that

$$\widehat{f}(\widehat{w}, v) = \widehat{\lambda}\widehat{g}(\widehat{w}, v) \text{ for every function } v \in V_J^0.$$

On the high probability event $\mathcal{R}_2 \subset \mathcal{R}_1$ such that $\widehat{\sigma}^2_{FZ} \sim 1$ (see Corollary 4.24), the form $\widehat{f}$ is positive-definite and symmetric. Consequently, on $\mathcal{R}_2$, the Eigenproblem 4.25 has J solutions $(\widehat{\lambda}_j, \widehat{w}_j)_{j=1,..,J}$ with $0 < \widehat{\lambda}_1 \leq \widehat{\lambda}_2 \leq ... \leq \widehat{\lambda}_J$.

Definition 4.26. For $j = 1, ..., J$ define $\psi_j^0 = \psi_j - \int_0^1 \psi_j(x)\widehat{\mu}_N(dx) \in V_J^0$. Let

$$\widehat{F}_{i,j} := \widehat{f}(\psi_i^0, \psi_j^0) = \widehat{f}(\psi_i, \psi_j) \quad \text{and} \quad \widehat{M}_{i,j} = \widehat{g}(\psi_i^0, \psi_j^0)$$

be the matrix representations of forms $\widehat{f}$ and $\widehat{g}$ on $V_J^0 \times V_J^0$ with respect to the algebraic basis $(\psi_j^0)_j$.

Arguing as in Gobet et al. [44, Lemma 6.1] we obtain that

$$\widehat{M}_{i,j} = \int_{\frac{i-1}{J}}^{\frac{i}{J}}\int_{\frac{j-1}{J}}^{\frac{j}{J}}\int_0^{y\wedge z} \widehat{\mu}_N(dx)\int_{y\vee z}^1 \widehat{\mu}_N(dx)dydz. \tag{4.27}$$

$\widehat{F}$ is a diagonal matrix with strictly positive diagonal entries, hence it is invertible. Eigenproblem 4.25 is equivalent to

$$\widehat{F}^{-1}\widehat{M}(\widehat{w}_{i,j})_j = \widehat{\lambda}_i^{-1}(\widehat{w}_{i,j})_j,$$

where $(\widehat{w}_{i,j})_{j=1,\dots,J}$ indicates the coefficient vector associated to the eigenfunction $\widehat{w}_i$, i.e.

$$\widehat{w}_i = \sum_{j=1}^{J} \widehat{w}_{i,j}\psi_j^N = \sum_{j=1}^{J} \widehat{w}_{i,j}\psi_j + \widehat{w}_{i,0},$$

with some $\widehat{w}_{i,0}$ such that $\widehat{w}_i \in V_J^0$. Since the matrix $\widehat{M}$ has all entries strictly positive, the matrix $\widehat{F}^{-1}\widehat{M}$ satisfies the conditions of the Perron–Frobenius theorem. Consequently, the eigenvector $(\widehat{w}_{1,j})_j$ can be chosen strictly positive, which corresponds to the monotonicity property of the eigenfunction $\widehat{w}_1$. In what follows, we will show that the Eigenproblem 4.25 is an approximation of the Eigenproblem 4.5 for forms $\widehat{l}$ and $\widehat{g}$, and deduce that the eigenfunction $\widehat{u}_1$ inherits the properties of $\widehat{w}_1$. Let $\|\cdot\|_{l^2}$ denote the standard Euclidean norm on $\mathbb{R}^J$.

Definition 4.27. Fix $\varepsilon > 0$. Theorem 4.17, Corollary 4.24, Theorem 4.19 and the regularity properties of the occupation density μ_1 ensure that there exist $\alpha = \alpha(\varepsilon) < 1/42$ and $C_{\varepsilon,\alpha}$ such that the set $\mathcal{R}_\alpha$ of paths contained in $\mathcal{L}_v$ that satisfy

(i) $\omega(\Delta) \le \Delta^{1/2-\alpha}$

(ii) for every $x \in (0,1)$ holds $\widehat{\sigma}^2_{FZ}(x) \sim 1$

(iii) $\|\widehat{\sigma}^2_{FZ} - \sigma^2\|_{L^2([1/J,1-1/J])} \le \Delta^{1/3-\alpha}$ and for $x \in [0,1/J] \cup [1-1/J,1]$ it holds $|\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)| \le \Delta^{1/33-\alpha}$

(iv) occupation density μ_1 is $1/2-\alpha$ Hölder continuous with Hölder norm bounded by $C_{\varepsilon,\alpha}$

is of high-probability, more precisely

$$\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}_\alpha) < \varepsilon.$$

Remark 4.28. By the assumption (iv) we have $\|\mu_1\|_\infty \lesssim 1$. Furthermore, arguing as in the proof of Lemma 4.18, we obtain

$$\Big|\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) - \int_{\frac{j-1}{J}}^{\frac{j}{J}} \mu_1(dx)\Big| \lesssim \omega(\Delta)\|\mu_1\|_\infty \lesssim \Delta^{1/2-\alpha}. \tag{4.28}$$

In particular, on $\mathcal{R}_\alpha$

$$\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) \sim \Delta^{1/3} \text{ holds for every } j = 1, \dots, J. \tag{4.29}$$

Finally, the assumption (iii) and Hölder regularity of σ^2 (4.14) imply that on the event $\mathcal{R}_\alpha$

$$|\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)| \lesssim \Delta^{1/6-\alpha} \text{ for all } x \in [1/J, 1-1/J]. \tag{4.30}$$

Proof of (4.30). Fix $j = 2, ..., J-1$ and $x \in [\frac{j-1}{J}, \frac{j}{J}]$. We have

$$\begin{aligned}|\widehat{\sigma}^2_{FZ,j} - \sigma^2(x)| &\le \Big|\widehat{\sigma}^2_{FZ,j} - J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(y)dy\Big| + \Big|\sigma^2(x) - J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(y)dy\Big| \\ &\lesssim J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{\sigma}^2_{FZ,j} - \sigma^2(y)|dy + \Delta^{1/6} \\ &\lesssim J^{1/2}\Big(\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{\sigma}^2_{FZ,j} - \sigma^2(y)|^2 dy\Big)^{\frac{1}{2}} + \Delta^{1/6} \\ &\lesssim J^{1/2}\|\widehat{\sigma}^2_{FZ} - \sigma^2\|_{L^2([1/J,1-1/J])} + \Delta^{1/6},\end{aligned}$$

where we used the $1/2-$Hölder regularity of σ^2 (4.14) and the Cauchy-Schwarz inequality to get the second and third inequalities respectively. Finally, since $J^{1/2} \sim \Delta^{-1/6}$, we conclude by Definition 4.27.(iii) that the claim holds. ☐

To bound the error between the solutions of the Eigenproblems 4.5 and 4.25 we need to establish uniform bounds on the spectral gap of the Eigenproblem 4.25.

Lemma 4.29. *On the event $\mathcal{R}_\alpha$ the eigenvalue $\widehat{\lambda}_1$ is uniformly bounded. Furthermore, the Eigenproblem 4.25 has a uniform spectral gap, i.e. $\widehat{\lambda}_1^{-1} - \widehat{\lambda}_2^{-1} \gtrsim 1$.*

Proof. Consider the generalized eigenvalue problem:

Eigenproblem 4.30. Find $(\lambda, w) \in \mathbb{R} \times V_J$ with $w \neq 0$ and $\int_0^1 w(x)\mu_1(x)dx = 0$ such that

$$\int_0^1 w'(x)v'(x)\sigma^2(x)\mu_1(x)dx = \lambda\int_0^1 w(x)v(x)\mu_1(x)dx, \qquad \text{for all } v \in \{v \in V_J : \int_0^1 v(x)\mu_1(x)dx = 0\}.$$

Eigenproblem 4.30 has J solutions, denoted by (λ_j, w_j) with $0 < \lambda_1 < \lambda_2 \le ... \le \lambda_J$. By Proposition A.8 we have $\lambda_1 \sim 1$ and $\lambda_1^{-1} - \lambda_2^{-1} \gtrsim 1$. Let M, F be $J \times J$ matrices corresponding to the Eigenproblem 4.30 tested with functions $(\psi_j^1)_{j=1,...,J}$, where $\psi_j^1 = \psi_J - \int_0^1 \psi_j\mu_1(x)dx$. As in the case of the data driven Eigenproblem 4.25, we have

$$\begin{aligned}M_{i,j} &= \int_{\frac{i-1}{J}}^{\frac{i}{J}}\int_{\frac{j-1}{J}}^{\frac{j}{J}}\int_0^{y\wedge z}\mu_1(x)dx\int_{y\vee z}^1 \mu_1(x)dxdydz, \\ F_{i,j} &= \begin{cases} 0 & : i \neq j \\ \int_{\frac{i-1}{J}}^{\frac{i}{J}} \sigma^2(x)\mu_1(x)dx & : i = j\end{cases}\end{aligned}$$

and

$$F^{-1}M(w_{i,j})_j = \lambda_i^{-1}(w_{i,j})_j.$$

From Weyl's theorem for symmetric eigenvalue problems it follows that

$$|\widehat{\lambda}_i^{-1} - \lambda_i^{-1}| \le \|F^{-1}M - \widehat{F}^{-1}\widehat{M}\|_{l^2}. \tag{4.31}$$

We will show that $\|F^{-1}M - \widehat{F}^{-1}\widehat{M}\|_{l^2} \lesssim \Delta^{1/6-\alpha}$. Then, the uniform bound on the eigenvalue $\widehat{\lambda}_1$ and the lower bound on the spectral gap will follow from the properties of the Eigenproblem 4.30. First, let us observe that

$$|\widehat{F}_{j,j} - F_{j,j}| = \Big| \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\sigma}^2_{FZ,j}\widehat{\mu}_N(dx) - \int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x)\mu_1(x)dx\Big| \lesssim$$

$$\lesssim \widehat{\sigma}^2_{FZ,j}\Big| \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) - \int_{\frac{j-1}{J}}^{\frac{j}{J}} \mu_1(x)dx\Big| + \int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{\sigma}^2_{FZ}(x) - \sigma^2(x)|\mu_1(x)dx.$$

On $\mathcal{R}_\alpha$, the first term of the above display is by (4.28) of order $\Delta^{1/2-\alpha}$. The second term is for $j = 2, ..., J-1$ of order $\Delta^{1/2-\alpha}$ by (4.30) and for $j = 1$ or J of order $\Delta^{1/3+1/33-\alpha}$ by Definition 4.27.(iii) directly. In any case, for Δ small enough it holds $\widehat{F}_{j,j} \sim F_{j,j} \sim \Delta^{1/3}$. Next, arguing as in the proof of Lemma 4.18, for any $i, j = 1, ..., J$, we obtain

$$|M_{i,j} - \widehat{M}_{i,j}| \lesssim J^{-2}\omega(\Delta)\|\mu_1\|_\infty \lesssim \Delta^{7/6-\alpha}.$$

Furthermore $M_{i,j}, \widehat{M}_{i,j} \lesssim \Delta^{2/3}$. Since F and $\widehat{F}$ are diagonal matrices, it follows that

$$|(F^{-1}M - \widehat{F}^{-1}\widehat{M})_{i,j}| \lesssim \Delta^{1/2-\alpha}.$$

Hence, $\|F^{-1}M - \widehat{F}^{-1}\widehat{M}\|^2_{l^2} \leq \sum_{i,j=1}^J (F^{-1}M - \widehat{F}^{-1}\widehat{M})^2_{i,j} \lesssim \Delta^{1/3-2\alpha}$. □

Proposition 4.31. *Choose the eigenfunction $\widehat{w}_1$ increasing and normalized so that $\|(\widehat{w}_{1,j})_j\|_{l^2} = J^{1/2}$ (i.e. $\|\widehat{w}_1'\|_{L^2} = 1$). On the event $\mathcal{R}_\alpha$, for any $\lfloor aJ\rfloor - 1 \leq j \leq \lceil bJ\rceil + 1$ and any $i = 1, ..., J$ we have*

$$1 \vee \widehat{w}_{1,i} \lesssim \widehat{w}_{1,j} \wedge 1, \tag{4.32}$$

Furthermore for j s.t. $J^{1/2} \leq j \leq J - J^{1/2}$

$$\Big|\frac{\widehat{w}_{1,j\pm1}}{\widehat{w}_{1,j}} - 1\Big| \lesssim \Delta^{1/6-\alpha}. \tag{4.33}$$

Proof. In the proof we will use standard techniques from the Perron-Frobenius theory of nonnegative matrices (cf. Minc [61, Chapter II]). In particular, we shall repeatedly use the following inequality Minc [61, Chapter II, Section 2.1, Eq. (7)]: for any $q_1, q_2, ..., q_n > 0$ and $p_1, p_2, ..., p_n \in \mathbb{R}$

$$\min_{i=1,...,n} \frac{p_i}{q_i} \leq \frac{p_1 + p_2 + ... + p_n}{q_1 + q_2 + ... + q_n} \leq \max_{i=1,...,n} \frac{p_i}{q_i}. \tag{4.34}$$

Step 1: ($\widehat{w}_{1,i} \lesssim 1$). Fix $1 \leq i \leq J$. By Definition 4.27.(ii), relation (4.29) and Lemma 4.29, on the event $\mathcal{R}_\alpha$, we have

$$J^{-1}\widehat{w}_{1,i} \sim \widehat{w}_{1,i}\widehat{\sigma}^2_{FZ,i}\int_{\frac{i-1}{J}}^{\frac{i}{J}} \mu_N(dx) = \widehat{f}(\widehat{w}_1, \psi_i) = \widehat{\lambda}_1\widehat{g}(\widehat{w}_1, \psi_i) \sim \widehat{g}(\widehat{w}_1, \psi_i) = \sum_{m=1}^J \widehat{M}_{i,m}\widehat{w}_{1,m}. \tag{4.35}$$

Hence, by the Cauchy-Schwarz inequality

$$\widehat{w}_{1,i} \lesssim J\Big(\sum_{m=1}^J M^2_{i,m}\Big)^{1/2}\Big(\sum_{m=1}^J \widehat{w}^2_{1,m}\Big)^{1/2} \lesssim 1,$$

where we used $M_{i,m} \leq J^{-2}$ and the normalization of $(\widehat{w}_{1,j})$.

Step 2: ($\widehat{w}_{1,i} \lesssim \widehat{w}_{1,j}$). Fix $\lfloor aJ \rfloor - 1 \leq j \leq \lceil bJ \rceil + 1$. On the event $\mathcal{R}_\alpha$, for any $1 \leq i \leq J$ the relation (4.35) together with the inequality (4.34) imply

$$\frac{\widehat{w}_{1,i}}{\widehat{w}_{1,j}} \sim \frac{\sum_{m=1}^J \widehat{M}_{i,m} \widehat{w}_{1,m}}{\sum_{m=1}^J \widehat{M}_{j,m} \widehat{w}_{1,m}} \lesssim \max_{m=1,\ldots,J} \frac{\widehat{M}_{i,m}}{\widehat{M}_{j,m}}. \tag{4.36}$$

We need to show that for arbitrary m $\widehat{M}_{i,m}/\widehat{M}_{j,m} \lesssim 1$ holds. Consider first the case $i < j$. Then by (4.27)

$$\begin{aligned}
\frac{\widehat{M}_{i,m}}{\widehat{M}_{j,m}} &= \frac{\int_{\frac{i-1}{J}}^{\frac{i}{J}} \int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_0^{y \wedge z} \widehat{\mu}_N(dx) \int_{y \vee z}^1 \widehat{\mu}_N(dx) dy dz}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_0^{y \wedge z} \widehat{\mu}_N(dx) \int_{y \vee z}^1 \widehat{\mu}_N(dx) dy dz} \\
&\leq \frac{\int_{\frac{i-1}{J}}^{\frac{i}{J}} \int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_0^{y \wedge \frac{j-1}{J}} \widehat{\mu}_N(dx) \int_{y \vee \frac{i-1}{J}}^1 \widehat{\mu}_N(dx) dy dz}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_0^{y \wedge \frac{j-1}{J}} \widehat{\mu}_N(dx) \int_{y \vee \frac{j}{J}}^1 \widehat{\mu}_N(dx) dy dz} \\
&= \frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) \int_{y \vee \frac{i-1}{J}}^1 \widehat{\mu}_N(dx) dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) \int_{y \vee \frac{j}{J}}^1 \widehat{\mu}_N(dx) dy},
\end{aligned}$$

where $f(y) = \int_0^{y \wedge \frac{j-1}{J}} \widehat{\mu}_N(dx)$. Consider $m > j$. For $y \in [\frac{m-1}{J}, \frac{m}{J}]$ holds $y = y \vee \frac{j}{J} = y \vee \frac{i-1}{J}$, hence the numerator and denominator are equal. Consider $m \leq j$. For $y \in [\frac{m-1}{J}, \frac{m}{J}]$ holds $y \vee \frac{j}{J} = \frac{j}{J}$. Hence, using (4.29), we obtain

$$\begin{aligned}
\frac{\widehat{M}_{i,m}}{\widehat{M}_{j,m}} &\leq \frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) \int_{y \vee \frac{i-1}{J}}^1 \widehat{\mu}_N(dx) dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) \int_{y \vee \frac{j}{J}}^1 \widehat{\mu}_N(dx) dy} \leq \frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}} f(y) dy \int_{\frac{j}{J}}^1 \widehat{\mu}_N(dx)} \\
&= \Big(\int_{\frac{j}{J}}^1 \widehat{\mu}_N(dx) \Big)^{-1} \sim \Big(1 - \frac{j}{J}\Big)^{-1} \lesssim 1.
\end{aligned}$$

We conclude that for $i < j$ and arbitrary m bound $\widehat{M}_{i,m}/\widehat{M}_{j,m} \lesssim 1$ holds. Proceeding analogously, we obtain the same claim for $i > j$. From (4.36) it follows that on the event $\mathcal{R}_\alpha$, for $\lfloor aJ \rfloor - 1 \leq j \leq \lceil bJ \rceil + 1$ and any $1 \leq i \leq J$, we have

$$\widehat{w}_{1,i} \lesssim \widehat{w}_{1,j}. \tag{4.37}$$

Step 3: ($1 \lesssim \widehat{w}_{1,j}$). Let $\widehat{w}_{1,j_0} = \min_{\lfloor aJ \rfloor - 1 \leq j \leq \lceil bJ \rceil + 1} \widehat{w}_{1,j}$. Inequality (4.37) implies

$$1 = \frac{1}{J} \sum_{i=1}^J \widehat{w}_{1,i}^2 \lesssim \widehat{w}_{1,j_0}^2.$$

Step 4: (proof of (4.33)). We will only show $\big|\frac{\widehat{w}_{1,j+1}}{\widehat{w}_{1,j}} - 1\big| \lesssim \Delta^{1/6-\alpha}$, the other bound can be obtained by a symmetric argument. First, note that from Definition 4.27.(iv) together with the inequality (4.28) it follows that

$$\Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\Big|\le\Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)-\int_{\frac{j}{J}}^{\frac{j+1}{J}}\mu_1(x)dx\Big|+$$
$$+\Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\mu_1(x)dx-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(x)dx\Big|+\Big|\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(x)dx-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\Big|$$
$$\lesssim\Delta^{1/2-\alpha}+\Delta^{1/3}\Delta^{(1/2-\alpha)/3}+\Delta^{1/2-\alpha}\lesssim\Delta^{1/2-\alpha}.$$

Hence, by (4.29)

$$\left|\frac{\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)}{\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}-1\right|=\left|\frac{\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}{\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}\right|\lesssim\Delta^{1/6-\alpha}.$$

Similarly, by the $1/2-$Hölder regularity of σ^2 and Definition 4.27.(ii) together with (4.30) we have

$$\Big|\frac{\widehat{\sigma}^2_{FZ,j+1}}{\widehat{\sigma}^2_{FZ,j}}-1\Big|\lesssim\Delta^{1/6-\alpha}.$$

Consequently, instead of $\frac{\widehat{w}_{1,j+1}}{\widehat{w}_{1,j}}$ we may consider

$$\frac{\widehat{w}_{1,j+1}\widehat{\sigma}^2_{FZ,j+1}\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)}{\widehat{w}_{1,j}\widehat{\sigma}^2_{FZ,j}\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}=\frac{\sum_{m=1}^J\widehat{M}_{j+1,m}\widehat{w}_{1,m}}{\sum_{m=1}^J\widehat{M}_{j,m}\widehat{w}_{1,m}}.$$

By the inequality (4.34)

$$\min_{m=1,\dots,J}\frac{\widehat{M}_{j+1,m}}{\widehat{M}_{j,m}}\le\frac{\sum_{m=1}^J\widehat{M}_{j+1,m}\widehat{w}_{1,m}}{\sum_{m=1}^J\widehat{M}_{j,m}\widehat{w}_{1,m}}\le\max_{m=1,\dots,J}\frac{\widehat{M}_{j+1,m}}{\widehat{M}_{j,m}}.$$

Thus, it is enough to show, that for any $m=1,\dots,J$ bound $\Big|\frac{\widehat{M}_{j+1,m}}{\widehat{M}_{j,m}}-1\Big|\lesssim\Delta^{1/6}$ holds.

$$\begin{aligned}\frac{\widehat{M}_{j+1,m}}{\widehat{M}_{j,m}}&=\frac{\int_{\frac{j}{J}}^{\frac{j+1}{J}}\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge z}\widehat{\mu}_N(dx)\int_{y\vee z}^1\widehat{\mu}_N(dx)dydz}{\int_{\frac{j-1}{J}}^{\frac{j}{J}}\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge z}\widehat{\mu}_N(dx)\int_{y\vee z}^1\widehat{\mu}_N(dx)dydz}\\&\le\frac{\int_{\frac{j}{J}}^{\frac{j+1}{J}}\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge\frac{j+1}{J}}\widehat{\mu}_N(dx)\int_{y\vee\frac{j}{J}}^1\widehat{\mu}_N(dx)dydz}{\int_{\frac{j-1}{J}}^{\frac{j}{J}}\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge\frac{j-1}{J}}\widehat{\mu}_N(dx)\int_{y\vee\frac{j}{J}}^1\widehat{\mu}_N(dx)dydz}\\&\le\frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge\frac{j+1}{J}}\widehat{\mu}_N(dx)f(y)dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge\frac{j-1}{J}}\widehat{\mu}_N(dx)f(y)dy}=1+\frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_{y\wedge\frac{j-1}{J}}^{y\wedge\frac{j+1}{J}}\widehat{\mu}_N(dx)f(y)dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}}\int_0^{y\wedge\frac{j-1}{J}}\widehat{\mu}_N(dx)f(y)dy},\end{aligned}$$

where $f(y) = \int_{y\vee\frac{j}{J}}^{1} \widehat{\mu}_N(dx)$. Consider $m \leq j-1$. For $y \in [\frac{m-1}{J}, \frac{m}{J}]$ we have $y = y \wedge \frac{j+1}{J} = y \wedge \frac{j-1}{J}$, hence the error term is zero. Consider $m \geq j$. For $y \in [\frac{m-1}{J}, \frac{m}{J}]$ we have $y \wedge \frac{j-1}{J} = \frac{j-1}{J}$. Consequently, using (4.29), we obtain that for $j \geq J^{1/2} \sim \Delta^{-1/6}$

$$\begin{aligned}\frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_{y\wedge\frac{j-1}{J}}^{y\wedge\frac{j+1}{J}} \widehat{\mu}_N(dx) f(y)dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_{0}^{y\wedge\frac{j-1}{J}} \widehat{\mu}_N(dx) f(y)dy} &\leq \frac{\int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_{\frac{j-1}{J}}^{\frac{j+1}{J}} \widehat{\mu}_N(dx) f(y)dy}{\int_{\frac{m-1}{J}}^{\frac{m}{J}} \int_{0}^{\frac{j-1}{J}} \widehat{\mu}_N(dx) f(y)dy} \\ &\sim \frac{2J^{-1}}{\frac{j-1}{J}} = \frac{2}{j-1} \lesssim \Delta^{1/6}.\end{aligned}$$

Finally, symmetric bound $1 - \frac{\widehat{M}_{j+1,m}}{\widehat{M}_{j,m}} \lesssim \Delta^{1/6}$ can be obtained by similar calculations. □

In previous proposition we have established uniform bounds on the eigenfunction $\widehat{w}_1$. Next, we show that $\widehat{w}_1$ is a good approximation of $\widehat{u}_1$.

Definition 4.32. Let $\widehat{L}$ be the matrix representation of the form $\widehat{l}$ with respect to the algebraic basis $(\psi_j^0)_j$ (see Definition 4.26), i.e.

$$\widehat{L}_{i,j} := \widehat{l}(\psi_i^0, \psi_j^0) = \widehat{l}(\psi_i, \psi_j).$$

On the event $\mathcal{R}_\alpha$, for Δ sufficiently small, the matrix $\widehat{L}$ is symmetric tridiagonal. We want to bound the error between the solutions of the generalized eigenproblems:

$$\widehat{M}(\widehat{w}_i) = \widehat{\lambda}_1^{-1} \widehat{F}(\widehat{w}_i) \quad \text{and} \quad \widehat{M}(\widehat{u}_i) = \widehat{\gamma}_1^{-1} \widehat{L}(\widehat{u}_i).$$

Lemma 4.33. *On the event $\mathcal{R}_\alpha$ holds*

$$\|\widehat{F} - \widehat{L}\|_{l^2} \lesssim \Delta^{1/2-3\alpha}. \tag{4.38}$$

Furthermore matrix $\widehat{L}$ is invertible and

$$\|\widehat{L}\|_{l^2}, \|\widehat{F}\|_{l^2}, \|\widehat{L}^{-1}\|_{l^2}^{-1}, \|\widehat{F}^{-1}\|_{l^2}^{-1} \sim \Delta^{1/3}.$$

Proof. Consider vector $(v_j)_j \in \mathbb{R}^J$ with $\|(v_j)_j\|_{l^2} = 1$ and the corresponding function $v = \sum_{j=1}^J v_j \psi_j^0(x) \in V_J^0$. Since

$$\begin{aligned}\|(\widehat{F} - \widehat{L})v\|_{l^2}^2 &= \sum_{j=1}^J |\widehat{f}(v, \psi_j) - \widehat{l}(v, \psi_j)|^2 \\ &= \sum_{j=1}^J \left(v_{j-1}\widehat{L}_{j-1,j} + v_j(\widehat{F}_{j,j} - \widehat{L}_{j,j}) + v_{j+1}\widehat{L}_{j+1,j}\right)^2,\end{aligned}$$

to obtain (4.38), we just have to argue that $\widehat{L}_{j-1,j}, |\widehat{F}_{j,j} - \widehat{L}_{j,j}|$ and $\widehat{L}_{j+1,j}$ are of order $\Delta^{1/2-3\alpha}$. By the definition of the forms $\widehat{l}$ and $\widehat{g}$ from the Eigenproblem 4.5

$$2|\widehat{L}_{j-1,j}| = \sum_{n=0}^{N-1} (\psi_{j-1}(X_{(n+1)\Delta}) - \psi_{j-1}(X_{n\Delta}))(\psi_j(X_{(n+1)\Delta}) - \psi_j(X_{n\Delta}))$$

$$
\begin{aligned}
&= \sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < \tfrac{j-1}{J})\mathbf{1}(X_{(n+1)\Delta} > \tfrac{j-1}{J})(\tfrac{j-1}{J} - X_{n\Delta})(X_{(n+1)\Delta} - \tfrac{j-1}{J}) \\
&\qquad + \sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} > \tfrac{j-1}{J})\mathbf{1}(X_{(n+1)\Delta} < \tfrac{j-1}{J})(\tfrac{j-1}{J} - X_{(n+1)\Delta})(X_{n\Delta} - \tfrac{j-1}{J}) \\
&\lesssim \sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < \tfrac{j-1}{J})\mathbf{1}(X_{(n+1)\Delta} > \tfrac{j-1}{J})(X_{(n+1)\Delta} - X_{n\Delta})^2 \\
&\qquad + \sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} > \tfrac{j-1}{J})\mathbf{1}(X_{(n+1)\Delta} < \tfrac{j-1}{J})(X_{(n+1)\Delta} - X_{n\Delta})^2.
\end{aligned}
$$

Moreover

$$
\begin{aligned}
|\widehat{F}_{j,j} - \widehat{L}_{j,j}| \leq \frac{1}{2}\sum_{n=0}^{N-1} |\mathbf{1}(X_{n\Delta} < \tfrac{j-1}{J}) - \mathbf{1}(X_{(n+1)\Delta} < \tfrac{j-1}{J})|(X_{(n+1)\Delta} - X_{n\Delta})^2 + \\
+ \frac{1}{2}\sum_{n=0}^{N-1} |\mathbf{1}(X_{n\Delta} < \tfrac{j}{J}) - \mathbf{1}(X_{(n+1)\Delta} < \tfrac{j}{J})|(X_{(n+1)\Delta} - X_{n\Delta})^2
\end{aligned}
$$

Hence, it suffices to show that for any $x \in (0,1)$

$$
\sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < x)\mathbf{1}(X_{(n+1)\Delta} > x)(X_{(n+1)\Delta} - X_{n\Delta})^2 \lesssim \Delta^{1/2-3\alpha}. \tag{4.39}
$$

By Definition 4.27.(i), on the event $\mathcal{R}_\alpha$, we have

$$
\begin{aligned}
&\sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < x)\mathbf{1}(X_{(n+1)\Delta} > x)(X_{(n+1)\Delta} - X_{n\Delta})^2 \\
&\qquad\qquad \leq \Delta^{-2\alpha}\frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}(|X_{n\Delta} - x| \leq \Delta^{1/2-\alpha}).
\end{aligned}
$$

Arguing as in the proof of Lemma 4.18, we finally obtain

$$
\frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}(|X_{n\Delta} - x| \leq \Delta^{1/2-\alpha}) \lesssim \int_{x-\Delta^{1/2-\alpha}}^{x+\Delta^{1/2-\alpha}} \mu_1(x)dx + \omega(\Delta)\|\mu_1\|_\infty \lesssim \Delta^{1/2-\alpha}.
$$

Since $\widehat{F}$ is a diagonal matrix with diagonal entries of order $\Delta^{1/3}$, we have $\|\widehat{F}\|_{l^2}, \|\widehat{F}^{-1}\|_{l^2}^{-1} \sim \Delta^{1/3}$. As argued above, on $\mathcal{R}_\alpha$, the upper and lower diagonal entries of $\widehat{L}$ are of order $\Delta^{1/2-3\alpha}$. Since for any $1 \leq j \leq J$ holds $|\widehat{L}_{j,j} - \widehat{F}_{j,j}| \lesssim \Delta^{1/2-3\alpha}$, matrix $\widehat{L}$ is diagonally dominant with diagonal entries of order $\Delta^{1/3}$. Hence it is invertible and $\|\widehat{L}\|_{l^2}, \|\widehat{L}^{-1}\|_{l^2}^{-1} \sim \Delta^{1/3}$. □

Lemma 4.34. *Eigenvectors $(\widehat{w}_{1,j})$, $(\widehat{u}_{1,j})$, normalized so that $\|\widehat{w}_1\|_{l^2} = \|\widehat{u}_1\|_{l^2} = J^{1/2}$, satisfy on $\mathcal{R}_\alpha$*

$$
\|(\widehat{w}_{1,j}) - (\widehat{u}_{1,j})\|_{l^2} \lesssim \Delta^{-1/3}\|(\widehat{F} - \widehat{L})\widehat{w}_1\|_{l^2}.
$$

Proof. Recall that $(\widehat{\lambda}_j, \widehat{w}_j)_j$ are the eigenpairs of the Eigenproblem 4.25, with $\|(\widehat{w}_j)\|_{l^2} = \sqrt{J}$.

[23, Theorem 26] implies that there exists an eigenpair $(\widehat{\lambda}_{j_0}, J^{-1/2}\widehat{w}_{j_0})$ such that

$$\begin{aligned}
|\widehat{\lambda}_{j_0}^{-1} - \widehat{\gamma}_1^{-1}| &\lesssim J^{-1/2}\|\widehat{F}^{-1}\|_{l^2}\|(\widehat{F}-\widehat{L})\widehat{w}_1\|_{l^2} \lesssim \|\widehat{F}^{-1}\|_{l^2}\|\widehat{F}-\widehat{L}\|_{l^2},\\
\|(\widehat{w}_{j_0,j}) - (\widehat{u}_{1,j})\|_{l^2} &\lesssim \delta^{-1}(\widehat{\lambda}_{j_0}^{-1})\|\widehat{F}^{-1}\|_{l^2}^{3/2}\|\widehat{F}\|_{l^2}^{1/2}\|(\widehat{F}-\widehat{L})\widehat{w}_1\|_{l^2},
\end{aligned}$$

where $\delta(\widehat{\lambda}_{j_0}^{-1})$ is the so called localizing distance, i.e. $\delta(\widehat{\lambda}_{j_0}^{-1}) = \min_{j\neq j_0}|\widehat{\lambda}_j^{-1} - \widehat{\gamma}_1^{-1}|$. From Lemma 4.33 we deduce

$$|\widehat{\lambda}_{j_0}^{-1} - \widehat{\gamma}_1^{-1}| \lesssim \Delta^{1/6-3\alpha}.$$

By Nakatsukasa [63, Theorem 8.3] for any $i = 1, ..., J$ we have

$$|\widehat{\lambda}_i^{-1} - \widehat{\gamma}_i^{-1}| \lesssim \|\widehat{L}^{-1}\|_{l^2}\|\widehat{\lambda}_i^{-1}(\widehat{F}-\widehat{L})\|_{l^2},$$

which together with Lemmas 4.29 and 4.33 imply

$$|\widehat{\lambda}_1^{-1} - \widehat{\gamma}_1^{-1}| \lesssim \Delta^{1/6-3\alpha}. \tag{4.40}$$

By Lemma 4.29 holds $|\widehat{\lambda}_1^{-1} - \widehat{\lambda}_2^{-1}| \gtrsim 1$, hence we must have $j_0 = 1$. Furthermore, from the same uniform lower bound on the spectral gap it follows

$$\delta(\widehat{\lambda}_{j_0}^{-1}) = \delta(\widehat{\lambda}_1^{-1}) \gtrsim 1.$$

Since by Lemma 4.33 we have $\|\widehat{F}^{-1}\|_{l^2}^{3/2}\|\widehat{F}\|_{l^2}^{1/2} \lesssim \Delta^{-1/3}$, we conclude that the claim holds. □

Proof of Proposition 4.20. Set $\varepsilon > 0$. By Remark 4.28 there exists α s.t. $\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}_\alpha) \le \varepsilon$. Set

$$\mathcal{R}_2 = \mathcal{R}_\alpha \cap \big\{\|\widehat{w}_1 - \widehat{u}_1\|_{l^2}^2 \le \Delta^{1/7-6\alpha}\big\}.$$

Step 1. We will show

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_\alpha}\cdot\|(\widehat{F}-\widehat{L})\widehat{w}_1\|_{l^2}^2\big]^{1/2} \lesssim \Delta^{5/12-3\alpha}. \tag{4.41}$$

In the proof of Lemma 4.33 we argued that for any $j = 1, ..., J$ holds

$$\widehat{l}(\psi_j, \psi_{j-1}), \widehat{l}(\psi_j, \psi_{j+1}), |\widehat{l}(\psi_j, \psi_j) - \widehat{f}(\psi_j, \psi_j)| \lesssim \Delta^{1/2-3\alpha}. \tag{4.42}$$

Hence, using the uniform upper bound $\forall i = 1, ..., J$ $\widehat{w}_{1,i} \lesssim 1$ from (4.32), we obtain that

$$|\widehat{l}(\widehat{w}_1, \psi_j) - \widehat{f}(\widehat{w}_1, \psi_j)| \lesssim \Delta^{1/2-3\alpha} \text{ for any } j = 1, ..., J. \tag{4.43}$$

We will use the regularity of the eigenfunction $\widehat{w}_1$ to strengthen (4.43). Consider $J^{1/2} \le j \le J - J^{1/2}$.By the tridiagonal structure of the form $\widehat{l}$ it holds

$$\begin{aligned}
&\widehat{l}(\widehat{w}_1, \psi_j) - \widehat{f}(\widehat{w}_1, \psi_j)\\
&\quad= \widehat{l}(\psi_{j-1}, \psi_j)(\widehat{w}_{1,j-1} - \widehat{w}_{1,j}) + \widehat{w}_{1,j}\widehat{l}(I, \psi_j) - \widehat{w}_{1,j}\widehat{f}(\psi_j, \psi_j) + \widehat{l}(\psi_{j+1}, \psi_j)(\widehat{w}_{1,j+1} - \widehat{w}_{1,j})\\
&\quad= \widehat{w}_{1,j}\Big[\widehat{l}(\psi_{j-1}, \psi_j)\Big(\frac{\widehat{w}_{1,j-1}}{\widehat{w}_{1,j}} - 1\Big) + \widehat{l}(I, \psi_j) - \widehat{f}(I, \psi_j) + \widehat{l}(\psi_{j+1}, \psi_j)\Big(\frac{\widehat{w}_{1,j+1}}{\widehat{w}_{1,j}} - 1\Big)\Big].
\end{aligned}$$

Hence, using the upper bound $\widehat{w}_{1,j} \lesssim 1$ from (4.32), we obtain that

$$|\widehat{l}(\widehat{w}_1, \psi_j) - \widehat{f}(\widehat{w}_1, \psi_j)| \lesssim \widehat{l}(\psi_{j-1}, \psi_j)\Big|\frac{\widehat{w}_{1,j-1}}{\widehat{w}_{1,j}} - 1\Big| + \\ + \big|\widehat{l}(I, \psi_j) - \widehat{f}(I, \psi_j)\big| + \widehat{l}(\psi_{j+1}, \psi_j)\Big|\frac{\widehat{w}_{1,j+1}}{\widehat{w}_{1,j}} - 1\Big|.$$

Inequalities (4.42) and (4.33) imply

$$\widehat{l}(\psi_{j-1}, \psi_j)\Big(\frac{\widehat{w}_{1,j-1}}{\widehat{w}_{1,j}} - 1\Big) + \widehat{l}(\psi_{j+1}, \psi_j)\Big(\frac{\widehat{w}_{1,j+1}}{\widehat{w}_{1,j}} - 1\Big)\Big) \lesssim \Delta^{2/3-4\alpha},$$

while, since $\mathcal{R}_\alpha \subset \mathcal{R}_1$, from Lemma 4.39 it follows

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_\alpha} \cdot \big|\widehat{l}(I, \psi_j) - \widehat{f}(I, \psi_j)\big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{2}{3}}.$$

We conclude that for $J^{1/2} \leq j \leq J - J^{1/2}$

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_\alpha} \cdot |\widehat{l}(\widehat{w}_1, \psi_j) - \widehat{f}(\widehat{w}_1, \psi_j)|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{2}{3}-4\alpha}. \tag{4.44}$$

Since $\alpha < \frac{1}{12}$ inequalities (4.43) and (4.44) imply

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_\alpha} \cdot \|(\widehat{F} - \widehat{L})\widehat{w}_1\|^2_{l^2}] = \sum_{j=1}^{J} \mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_\alpha} \cdot |\widehat{f}(\widehat{w}_1, \psi_j) - \widehat{l}(\widehat{w}_1, \psi_j)|^2] \\ \lesssim J^{1/2}\Delta^{1-6\alpha} + J\Delta^{4/3-8\alpha} \lesssim \Delta^{5/6-6\alpha}.$$

Step 2. $\mathcal{R}_2$ is a high probability event. Indeed, inequality (4.41) and Lemma 4.34 imply

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_\alpha} \cdot \|\widehat{w}_1 - \widehat{u}_1\|^2_{l^2}]^{1/2} \lesssim \Delta^{1/12-3\alpha}.$$

Hence, by Markov's inequality,

$$\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}_2) \leq 2\varepsilon + \Delta^{-1/7+6\alpha}\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_\alpha} \cdot \|\widehat{w}_1 - \widehat{u}_1\|^2_{l^2}] \leq 2\varepsilon + C_\alpha\Delta^{1/6-1/7} \leq 3\varepsilon$$

for Δ sufficiently small.

Step 3. *O*n the event $\mathcal{R}_2$ holds

$$\max_{i=1,\dots,J} |\widehat{w}_{1,i} - \widehat{u}_{1,i}|^2 \leq \sum_{i=1}^{J} |\widehat{w}_{1,i} - \widehat{u}_{1,i}|^2 = \|\widehat{w}_1 - \widehat{u}_1\|^2_{l^2} \lesssim \Delta^{1/7-6\alpha}.$$

Since $\alpha < 1/42$ the eigenvector $(\widehat{u}_{1,j})$ inherits the uniform bounds of the eigenvector $(\widehat{w}_{1,j})$. In particular, for any $j = \lfloor aJ \rfloor - 1, ..., \lceil bJ \rceil + 1$, we have

$$\widehat{u}_{1,j} \sim 1.$$

Moreover, since for any $j = 1, ..., J$ holds $\widehat{w}_{1,j} > 0$, we deduce that

$$\sum_{j=1}^{J} \widehat{u}_{1,j}^2 \mathbf{1}(\widehat{u}_{1,j} < 0) \leq \|\widehat{w}_1 - \widehat{u}_1\|_{l^2}^2 \lesssim 1.$$

Finally, note that on the event $\mathcal{R}_2$ the eigenvalue $\widehat{\gamma}_1 \sim 1$ since on $\mathcal{R}_\alpha$, by (4.40), holds $|\widehat{\lambda}_1^{-1} - \widehat{\gamma}_1^{-1}| \lesssim \Delta^{1/6-3\alpha} \lesssim 1$ and $\widehat{\lambda}_1^{-1} \sim 1$ by Lemma 4.29. □

4.3.5. Proof of Theorem 4.8

As announced in Section 4.2.2, we will bound the approximation error of the spectral estimator and the time symmetric Florens-Zmirou estimator by the difference of forms $\widehat{f}$ and $\widehat{l}$.

Lemma 4.35. *On the high probability event $\mathcal{R}_2$ from Proposition 4.20 holds*

$$\|\widetilde{\sigma}_S^2 - \widehat{\sigma}_{FZ}^2\|_{L^1([a,b])} \lesssim \sum_{j=\lfloor aJ \rfloor}^{\lceil bJ \rceil} |\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)|.$$

Proof. From representations (4.22) and (4.10) it follows that

$$\|\widetilde{\sigma}_S^2 - \widehat{\sigma}_{FZ}^2\|_{L^1([a,b])} = \frac{1}{J} \sum_{j=\lfloor aJ \rfloor}^{\lceil bJ \rceil} |\widetilde{\sigma}_{S,j}^2 - \widehat{\sigma}_{FZ,j}^2| \lesssim \frac{1}{J} \sum_{j=\lfloor aJ \rfloor}^{\lceil bJ \rceil} \frac{|\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)|}{\widehat{u}_{1,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)}.$$

By Proposition 4.20, for $j = \lfloor aJ \rfloor - 1 \leq j \leq \lceil bJ \rceil + 1$, we have $\widehat{u}_{1,j} \sim 1$. Since, by Lemma 4.18, $J \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) \sim 1$, we conclude that the claim holds. □

Proposition 4.36. *For every function $v \in V_J^0$ and any $j = 1, ..., J$ we have*

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot |\widehat{f}(v, \psi_j) - \widehat{l}(v, \psi_j)|^2\big]^{\frac{1}{2}} \lesssim (|v_{j-1}|^2 + |v_j|^2 + |v_{j+1}|^2)^{\frac{1}{2}} \Delta^{\frac{1}{2}},$$

where v corresponds to the vector $(v_j)_{j=1,...,J}$ and $v_0, v_{J+1} = 0$.

Proof. First, note that since for $i \neq j$ holds $\widehat{f}(\psi_i, \psi_j) = 0$ we have $\widehat{f}(v, \psi_j) = v_j \widehat{f}(\psi_j, \psi_j)$. Moreover, on the event $\mathcal{R}_1$, for Δ sufficiently small, the increments of the process X are smaller than J^{-1}. Hence, for $|i - j| > 1$, holds $\widehat{l}(\psi_i, \psi_j) = 0$. Linearity implies

$$\widehat{l}(v, \psi_j) = v_{j-1}\widehat{l}(\psi_{j-1}, \psi_j) + v_j \widehat{l}(\psi_j, \psi_j) + v_{j+1}\widehat{l}(\psi_{j+1}, \psi_j). \tag{4.45}$$

Consequently, it is sufficient to show that

$$\begin{aligned} &\mathbb{E}_{\sigma,b}\big[|\widehat{l}(\psi_{j-1}, \psi_j)|^2\big]^{\frac{1}{2}} + \mathbb{E}_{\sigma,b}\big[|\widehat{f}(\psi_j, \psi_j) - \widehat{l}(\psi_j, \psi_j)|^2\big]^{\frac{1}{2}} + \\ &\qquad + \mathbb{E}_{\sigma,b}\big[|\widehat{l}(\psi_j, \psi_{j-1})|^2\big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{2}}. \end{aligned} \tag{4.46}$$

Decomposing the terms above like in Lemma 4.33, we obtain that (4.46) follows from Theorem 4.41. □

We are now able to prove the suboptimal rate $\Delta^{1/6}$ for the root mean squared $L^2([a,b])$ error of the spectral estimator $\widetilde{\sigma}_S$.

Proposition 4.37. *For every $\varepsilon > 0$ and Δ sufficiently small, there exists an event $\mathcal{R}_3 = \mathcal{R}_3(\varepsilon) \subseteq \mathcal{R}_2$, with $\mathbb{P}_{\sigma,b}(\mathcal{L}_v \setminus \mathcal{R}_3) \leq \varepsilon$, such that for every $x \in (a,b)$*

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_3} \cdot |\widetilde{\sigma}_S^2(x) - \sigma^2(x)|^2\big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{6}}. \tag{4.47}$$

Furthermore, on $\mathcal{R}_3$, for every $\lfloor aJ \rfloor \leq j \leq \lceil bJ \rceil$ we have

$$\widetilde{\sigma}_{S,j}^2 \sim 1, \tag{4.48}$$

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widehat{u}_{1,j\pm1}}{\widehat{u}_{1,j}} - 1\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{6}}. \tag{4.49}$$

Remark 4.38. Given the uniform lower bound on the derivative $\widehat{u}_{1,j}$, and since $\Delta^{1/6} \sim J^{-1/2}$, inequality (4.49) can be reformulated as

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widehat{u}_1'(\frac{j}{J} \pm \frac{1}{J}) - \widehat{u}_1'(\frac{j}{J})}{J^{-1/2}}\Big|^2\Big]^{\frac{1}{2}} \lesssim 1.$$

By means of Markov's inequality the latter can be interpreted as almost $1/2-$Hölder regularity of $\widehat{u}_1'$. In that sense Proposition 4.37 is a discrete time equivalent of Proposition A.5, which states that the derivatives of the eigenfunctions inherit the regularity of the design density, in the high-frequency case the regularity of the local time.

Proof of Proposition 4.37. Fix $\varepsilon > 0$. Let $\mathcal{R}_2$ be the high probability event introduced in Proposition 4.20. On $\mathcal{R}_2$, we choose the eigenfunction $\widehat{u}_1$ s.t.

$$\sum_{j=1}^{J} \widehat{u}_{1,j}^2 = J \quad \text{and} \quad \widehat{u}_{1,j} \sim 1 \quad \text{for every} \quad \lfloor aJ \rfloor - 1 \leq j \leq \lceil bJ \rceil + 1. \tag{4.50}$$

Step 1. Proof of (4.48). On the event $\mathcal{R}_1$, for Δ sufficiently small, using the representation (4.22) together with (4.45) and (4.50) we obtain that

$$\frac{\widehat{l}(\psi_j,\psi_j)}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)} \lesssim \widetilde{\sigma}_{S,j}^2 \lesssim \frac{\widehat{l}(\psi_{j-1}+\psi_j+\psi_{j+1},\psi_j)}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)} \tag{4.51}$$

holds for every $\lfloor aJ \rfloor \leq j \leq \lceil bJ \rceil$. Since

$$\widehat{l}(\psi_{j-1}+\psi_j+\psi_{j+1},\psi_j) \lesssim \sum_{n=0}^{N-1} (\mathbf{1}_j(X_{n\Delta}) + \mathbf{1}_j(X_{(n+1)\Delta}))(X_{(n+1)\Delta} - X_{n\Delta})^2,$$

we deduce that $\widetilde{\sigma}_{S,j}^2 \lesssim \widehat{\sigma}_{FZ,j}^2$. Furthermore, since on $\mathcal{R}_2$ holds

$$\widehat{l}(\psi_j,\psi_j) \geq \frac{1}{2} \sum_{n=0}^{N-1} \mathbf{1}(\tfrac{j-1}{J} + \Delta^{5/11} \leq X_{n\Delta} \leq \tfrac{j-1}{J} - \Delta^{5/11})(X_{(n+1)\Delta} - X_{n\Delta})^2, \tag{4.52}$$

the spectral estimator can be bounded from below by a time symmetric Florens-Zmirou estimator with bandwidth $\frac{1}{2}\Delta^{1/3} - \Delta^{5/11} \sim \Delta^{1/3}$. Arguing as in Corollary 4.24, we deduce that there exists a high probability event $\mathcal{R}_{3,1}$, such that on $\mathcal{R}_{3,1}$, bound $\widetilde{\sigma}^2_S(x) \gtrsim 1$ holds for any $x \in (a,b)$. Set

$$\mathcal{R}_3 = \mathcal{R}_2 \cap \mathcal{R}_{3,1}.$$

Step 2. Proof of (4.47). Fix $x \in (a,b)$ and chose j s.t. $\frac{j-1}{J} \le x < \frac{j}{J}$. Representations (4.22) and (4.10), together with Lemma 4.18, imply

$$|\widetilde{\sigma}^2_{S,j} - \widehat{\sigma}^2_{FZ,j}| \lesssim \Delta^{-1/3} |\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)|.$$

Hence, from Proposition 4.36 and (4.50) it follows that

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_3} \cdot |\widetilde{\sigma}^2_S(x) - \widehat{\sigma}^2_{FZ}(x)|^2\big]^{\frac{1}{2}} \lesssim \Delta^{1/6}.$$

By Theorem 4.19 and Hölder regularity of σ^2

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{R}_1} \cdot \|\sigma^2 - \widehat{\sigma}^2_{FZ}\|^2_\infty\big]^{\frac{1}{2}} \lesssim \Delta^{1/6}.$$

By the triangle inequality we conclude that (4.47) holds.

Step 3. Proof of (4.49). Set $\lfloor aJ \rfloor \le j \le \lceil bJ \rceil$. We will only prove

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widehat{u}_{1,j+1}}{\widehat{u}_{1,j}} - 1\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{6}}, \tag{4.53}$$

as the symmetric bound on the second moment of $\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widehat{u}_{1,j-1}}{\widehat{u}_{1,j}} - 1\Big|$ can be obtained analogously. The general idea of the proof is similar to the proof of (4.33) in Proposition 4.31. First, we will show that (4.53) follows from

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widehat{u}_{1,j+1}\widetilde{\sigma}^2_{S,j+1} \int_{\frac{j}{J}}^{\frac{j+1}{J}} \widehat{\mu}_N(dx)}{\widehat{u}_{1,j}\widetilde{\sigma}^2_{S,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)} - 1\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{1}{6}}. \tag{4.54}$$

To that purpose, by the triangle inequality and since on $\mathcal{R}_3$ the derivatives $\widehat{u}_{1,j}, \widehat{u}_{1,j+1} \sim 1$, we have to argue that

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\frac{\widetilde{\sigma}^2_{S,j+1} \int_{\frac{j}{J}}^{\frac{j+1}{J}} \widehat{\mu}_N(dx)}{\widetilde{\sigma}^2_{S,j} \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)} - 1\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{1/6}. \tag{4.55}$$

Step 3.1. Proof of (4.55). By Lemma 4.18 holds $J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx), J\int_{\frac{j}{J}}^{\frac{j+1}{J}} \widehat{\mu}_N(dx) \sim 1$. We defined above the event $\mathcal{R}_3$ so that $\widetilde{\sigma}^2_{S,j}, \widetilde{\sigma}^2_{S,j+1} \sim 1$. Hence, to prove (4.55), it suffices to show

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot |\widetilde{\sigma}^2_{S,j+1} - \widetilde{\sigma}^2_{S,j}|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{1/6} \tag{4.56}$$

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3} \cdot \Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}} \widehat{\mu}_N(dx) - \int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx)\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{1/2}. \tag{4.57}$$

(4.56) follows from (4.47) and 1/2 Hölder regularity of σ^2. Indeed

$$\begin{aligned}
\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3}\cdot\big|\widetilde{\sigma}^2_{S,j+1}-\widetilde{\sigma}^2_{S,j}\big|^2\Big]^{\frac{1}{2}} &\lesssim \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3}\cdot\big|\widetilde{\sigma}^2_{S,j+1}-\sigma^2(\tfrac{j+1/2}{J})\big|^2\Big]^{\frac{1}{2}}+\\
&+\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3}\cdot\big|\sigma^2(\tfrac{j+1/2}{J})-\sigma^2(\tfrac{j-1/2}{J})\big|^2\Big]^{\frac{1}{2}}+\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_3}\cdot\big|\sigma^2(\tfrac{j-1/2}{J})-\widetilde{\sigma}^2_{S,j}\big|^2\Big]^{\frac{1}{2}}\\
&\lesssim \Delta^{1/6}.
\end{aligned}$$

To prove (4.57) let

$$\begin{aligned}
\Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\Big| &\le \Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)-\int_{\frac{j}{J}}^{\frac{j+1}{J}}\mu_1(x)dx\Big|+\\
&+\Big|\int_{\frac{j}{J}}^{\frac{j+1}{J}}\mu_1(x)dx-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(x)dx\Big|+\Big|\int_{\frac{j-1}{J}}^{\frac{j}{J}}\mu_1(x)dx-\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\Big|\\
&:= E_1+E_2+E_3.
\end{aligned}$$

By [Supplement A, Theorem 11] we have

$$\mathbb{E}_{\sigma,b}[E_1^2+E_3^2]^{\frac{1}{2}}\lesssim\Delta^{2/3},$$

while the Cauchy-Schwarz inequality, together with [Supplement A, Theorem 8] yield

$$\begin{aligned}
\mathbb{E}_{\sigma,b}[E_2^2]^{\frac{1}{2}} &= \mathbb{E}_{\sigma,b}\Big[\Big|\int_0^{J^{-1}}\mu_1(\tfrac{j}{J}+x)-\mu_1(\tfrac{j-1}{J}+x)dx\Big|^2\Big]^{\frac{1}{2}}\\
&\le\Big[\frac{1}{J}\int_0^{J^{-1}}\mathbb{E}_{\sigma,b}\big[|\mu_1(\tfrac{j}{J}+x)-\mu_1(\tfrac{j-1}{J}+x)|^2\big]dx\Big]^{\frac{1}{2}}\lesssim\Delta^{\frac{1}{2}}.
\end{aligned}$$

Step 3.2. Proof of (4.54). The representation (4.22), together with the eigenpair property of $(\widehat{\gamma}_1,\widehat{u}_1)$, imply that

$$\frac{\widehat{u}_{1,j+1}\widetilde{\sigma}^2_{S,j+1}\int_{\frac{j}{J}}^{\frac{j+1}{J}}\widehat{\mu}_N(dx)}{\widehat{u}_{1,j}\widetilde{\sigma}^2_{S,j}\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)}=\frac{\widehat{l}(\widehat{u}_1,\psi_{j+1})}{\widehat{l}(\widehat{u}_1,\psi_j)}=\frac{\widehat{g}(\widehat{u}_1,\psi_{j+1})}{\widehat{g}(\widehat{u}_1,\psi_j)}.$$

In what follows we want to apply methods from the Perron-Frobenius theory for nonnegative matrices. To that purpose recall the definition of matrix $\widehat{M}$ from Section 4.3.4 Eq. (4.27). We have

$$\frac{\widehat{g}(\widehat{u}_1,\psi_{j+1})}{\widehat{g}(\widehat{u}_1,\psi_j)}=\frac{\sum_{m=1}^J\widehat{M}_{m,j+1}\widehat{u}_{1,m}}{\sum_{m=1}^J\widehat{M}_{m,j}\widehat{u}_{1,m}}.$$

To bound the above ratio we would like to proceed as in the proof of inequality (4.33) in Proposition 4.31. Unfortunately, we cannot, as we do not know if the vector of derivatives $(\widehat{u}_{1,j})$ is positive. Still, using the inequality (4.34) and arguing as in the proof of (4.33), we obtain that

$$\Big|\frac{\sum_{m=1}^J\widehat{M}_{m,j+1}\widehat{u}_{1,m}\mathbf{1}(\widehat{u}_{1,m}>0)}{\sum_{m=1}^J\widehat{M}_{m,j}\widehat{u}_{1,m}\mathbf{1}(\widehat{u}_{1,m}>0)}-1\Big|\lesssim\Delta^{1/6}.$$

To finish the proof we need to show that the possible error due to the negative derivative terms is small enough. On the event $\mathcal{R}_2$ we have

$$\widehat{g}(\widehat{u}_1,\psi_j)=\widehat{\gamma}_1^{-1}\widehat{l}(\widehat{u}_1,\psi_j)\sim\widehat{l}(\widehat{u}_1,\psi_j)\geq\widehat{u}_{1,j}\widehat{l}(\psi_j,\psi_j)\sim\widehat{l}(\psi_j,\psi_j).$$

Furthermore, on the event $\mathcal{R}_3$ we have $\widehat{l}(\psi_j,\psi_j)\gtrsim\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)$; indeed we defined $\mathcal{R}_3$ such that the left hand side of (4.51) has a uniform lower bound. Thus, by Lemma 4.18

$$\sum_{m=1}^{J}\widehat{M}_{m,j}\widehat{u}_{1,m}=\widehat{g}(\widehat{u}_1,\psi_j)\gtrsim\int_{\frac{j-1}{J}}^{\frac{j}{J}}\widehat{\mu}_N(dx)\gtrsim\Delta^{1/3}.$$

Consequently, we need to show that

$$\sum_{m=1}^{J}(\widehat{M}_{m,j}+\widehat{M}_{m,j+1})|\widehat{u}_{1,m}|\mathbf{1}(\widehat{u}_{1,m}\leq 0)\lesssim\Delta^{\frac{1}{2}}.$$

From (4.27) follows $\widehat{M}_{i,j}\lesssim J^{-2}$. By the Cauchy-Schwarz inequality and Proposition 4.20

$$\sum_{m=1}^{J}(\widehat{M}_{m,j}+\widehat{M}_{m,j+1})|\widehat{u}_{1,m}|\mathbf{1}(\widehat{u}_{1,m}\leq 0)\lesssim$$
$$\lesssim J^{-3/2}\Big(\sum_{m=1}^{J}|\widehat{u}_{1,m}|^2\mathbf{1}(\widehat{u}_{1,m}\leq 0)\Big)^{\frac{1}{2}}\lesssim\Delta^{\frac{1}{2}}.\qquad\square$$

To obtain the suboptimal rate $\Delta^{1/6}$ we only used uniform bounds on the derivatives vector $(\widehat{u}_{1,j})_j$ together with the general error bound from Proposition 4.36. Having established the regularity of the eigenfunction $\widehat{u}_1$, we are now able to argue that the error $\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{R}_3}\cdot|\widehat{l}(\widehat{u}_1,\psi_j)-\widehat{f}(\widehat{u}_1,\psi_j)|]$ is at most of order $\Delta^{2/3}$.

Lemma 4.39. *Denote $I(x)=x-c_0$, with c_0 such that $I\in V_J^0$. For Δ sufficiently small, for every $j=1,...,J$, it holds*

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1}\cdot|\widehat{f}(I,\psi_j)-\widehat{l}(I,\psi_j)|^2\Big]^{\frac{1}{2}}\lesssim\Delta^{2/3}.\tag{4.58}$$

Proof. We will reduce (4.58) to the term bounded in Theorem 4.42. By definition of the forms $\widehat{l}$, $\widehat{f}$ and the representation (4.10) it holds

$$\begin{aligned}\widehat{l}(I,\psi_j)&=\frac{1}{2}\sum_{n=0}^{N-1}(X_{(n+1)\Delta}-X_{n\Delta})(\psi_j(X_{(n+1)\Delta})-\psi_j(X_{n\Delta})),\\ \widehat{f}(I,\psi_j)&=\frac{1}{4}\sum_{n=0}^{N-1}(\mathbf{1}_j(X_{n\Delta})+\mathbf{1}_j(X_{(n+1)\Delta}))(X_{(n+1)\Delta}-X_{n\Delta})^2.\end{aligned}$$

We will analyze the error contribution of a single summand. When $X_{n\Delta},X_{(n+1)\Delta}\in[\frac{j-1}{J},\frac{j}{J}]$ both forms contribute by $\frac{1}{2}(X_{(n+1)\Delta}-X_{n\Delta})^2$, hence cancel perfectly. When $X_{n\Delta},X_{(n+1)\Delta}$ do not belong to $[\frac{j-1}{J},\frac{j}{J}]$ neither of the forms contribute. Since on $\mathcal{R}_1$, for Δ sufficiently small, the increment $|X_{(n+1)\Delta}-X_{n\Delta}|\leq 1/J$ we deduce that the overall error $|\widehat{f}(I,\psi_j)-\widehat{l}(I,\psi_j)|$ is due only to summands with the increment $X_{n\Delta}, X_{(n+1)\Delta}$ crossing the boundary of $[\frac{j-1}{J}, \frac{j}{J}]$. In such case the form $\widehat{f}$ contributes by $\frac{1}{4}(X_{(n+1)\Delta} - X_{n\Delta})^2$, while $\widehat{l}$ by $\frac{1}{2}(X_{(n+1)\Delta} - X_{n\Delta})\beta$, where

$$\beta = \operatorname{sgn}(X_{(n+1)\Delta} - X_{n\Delta}) \cdot \text{length}\left([X_{n\Delta}, X_{(n+1)\Delta}] \cap [\tfrac{j-1}{J}, \tfrac{j}{J}]\right).$$

Let $\gamma = X_{(n+1)\Delta} - X_{n\Delta} - \beta$. The contribution of a single boundary crossing summand equals

$$\frac{1}{4}(X_{(n+1)\Delta} - X_{n\Delta})^2 - \frac{1}{2}(X_{(n+1)\Delta} - X_{n\Delta})\beta = \frac{1}{4}(\beta + \gamma)(\gamma - \beta) = \frac{\gamma^2 - \beta^2}{4}.$$

Considering all four possible crossing configurations, we obtain that

$$\begin{aligned}\widehat{f}(I, \psi_j) - \widehat{l}(I, \psi_j) = &\sum_{n=0}^{N-1} \left(\mathbf{1}_{(\frac{j}{J},1]}(X_{(n+1)\Delta}) - \mathbf{1}_{(\frac{j}{J},1]}(X_{n\Delta})\right) \cdot \\ &\qquad \cdot \left((X_{(n+1)\Delta} - \tfrac{j}{J})^2 - (X_{n\Delta} - \tfrac{j}{J})^2\right) \\ + &\sum_{n=0}^{N-1} \left(\mathbf{1}_{(\frac{j-1}{J},1]}(X_{(n+1)\Delta}) - \mathbf{1}_{(\frac{j-1}{J},1]}(X_{n\Delta})\right) \cdot \\ &\qquad \cdot \left((X_{(n+1)\Delta} - \tfrac{j-1}{J})^2 - (X_{n\Delta} - \tfrac{j-1}{J})^2\right).\end{aligned}$$

Thus, (4.58) indeed follows from Theorem 4.42. □

Proof of Theorem 4.8. Set $\varepsilon > 0$. Let $\mathcal{R}_3$ be the high probability event introduced in Proposition 4.37. In view of Remark 4.22, it is enough to prove the claim for the estimator $\widetilde{\sigma}_S^2$. By Lemma 4.35 and since $J \sim \Delta^{-1/3}$, it is sufficient to show that for any $\lfloor aJ \rfloor \le j \le \lceil bJ \rceil$ holds

$$\mathbb{E}_{\sigma,b}\left[\mathbf{1}_{R_3} \cdot |\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j)|\right] \lesssim \Delta^{2/3}.$$

By Definition 4.14 holds $\widehat{f}(\widehat{u}_1, \psi_j) = \widehat{u}_{1,j}\widehat{f}(\psi_j, \psi_j) = \widehat{u}_{1,j}\widehat{f}(I, \psi_j)$. Since on the event $\mathcal{R}_3$, for Δ sufficiently small, the increments $|X_{(n+1)\Delta} - X_{n\Delta}| \le J^{-1}$, we have

$$\begin{aligned}\widehat{l}(\widehat{u}_1, \psi_j) &= \widehat{u}_{1,j-1}\widehat{l}(\psi_j, \psi_{j-1}) + \widehat{u}_{1,j}\widehat{l}(\psi_j, \psi_j) + \widehat{u}_{1,j+1}\widehat{l}(\psi_j, \psi_{j+1}), \\ \widehat{l}(I, \psi_j) &= \widehat{l}(\psi_j, \psi_{j-1}) + \widehat{l}(\psi_j, \psi_j) + \widehat{l}(\psi_j, \psi_{j+1}).\end{aligned}$$

Consequently, since by Proposition 4.20 $\widehat{u}_{1,j} \sim 1$, we deduce that

$$\begin{aligned}\widehat{l}(\widehat{u}_1, \psi_j) - \widehat{f}(\widehat{u}_1, \psi_j) \sim &\widehat{l}(\psi_j, \psi_{j-1})\Big(\frac{\widehat{u}_{1,j-1}}{\widehat{u}_{1,j}} - 1\Big) + \widehat{l}(I, \psi_j) - \widehat{f}(I, \psi_j) + \\ &+ \widehat{l}(\psi_j, \psi_{j+1})\Big(\frac{\widehat{u}_{1,j+1}}{\widehat{u}_{1,j}} - 1\Big).\end{aligned} \tag{4.59}$$

By the Cauchy-Schwarz inequality together with Proposition 4.36 and the inequality (4.49) we can uniformly bound the mean absolute value of the first and third term by $\Delta^{2/3}$. Since $\mathcal{R}_3 \subset \mathcal{R}_1$ the mean absolute value of the second term is bounded in Lemma 4.39. □

4.3.6. Technical results

We devote this chapter to the proof of two technical results that provide us with control over, properly rescaled, mean number of crossing of a given level α.

Definition 4.40. For $\alpha \in (0,1)$ and $n = 0, ..., N-1$ define

$$\chi(n,\alpha) = \mathbf{1}_{[0,\alpha)}(X_{(n+1)\Delta}) - \mathbf{1}_{[0,\alpha)}(X_{n\Delta}).$$

The random variable χ codifies the event of the increment $X_{n\Delta}, X_{(n+1)\Delta}$ crossing the level α. The sign of χ contains information about the direction of the crossing. Since

$$|\chi(n,\alpha)| \leq \mathbf{1}(|X_{n\Delta} - \alpha| \leq \omega(\Delta)),$$

arguing as in the proof of Lemma (4.18) we can show that

$$\frac{1}{N}\sum_{n=0}^{N-1} |\chi(n,\alpha)| \leq 4\omega(\Delta)\mu_1.$$

Consequently, Theorem 4.17 implies that the mean number of crossings, rescaled by the sample size, can be upper bounded by $\Delta^{1/2}\log(\Delta)$. Keeping in mind that $(X_{(n+1)\Delta} - X_{n\Delta})^2$ is of the order $\Delta = 1/N$, the next result is a refinement of the bound above.

Theorem 4.41. For every $\alpha \in (0,1)$ we have

$$\mathbb{E}_{\sigma,b}\Big[\Big(\sum_{n=0}^{N-1} |\chi(n,\alpha)|(X_{(n+1)\Delta} - X_{n\Delta})^2\Big)^2\Big]^{\frac{1}{2}} \lesssim \Delta^{1/2}.$$

Proof. Fix $\alpha \in (0,1)$. Since$|\chi(n,\alpha)| = 1$ if and only if the increment $(X_{n\Delta}, X_{(n+1)\Delta})$ crosses the level α, the claim is equivalent to the inequalities:

$$\begin{aligned}
\mathbb{E}_{\sigma,b}\Big[\Big(\sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < \alpha)\mathbf{1}(X_{(n+1)\Delta} > \alpha)(X_{(n+1)\Delta} - X_{n\Delta})^2\Big)^2\Big]^{\frac{1}{2}} &\lesssim \Delta^{1/2},\\
\mathbb{E}_{\sigma,b}\Big[\Big(\sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} > \alpha)\mathbf{1}(X_{(n+1)\Delta} < \alpha)(X_{(n+1)\Delta} - X_{n\Delta})^2\Big)^2\Big]^{\frac{1}{2}} &\lesssim \Delta^{1/2}.
\end{aligned}$$

Below, we only prove the first inequality. The second one can be obtained in a similar way or by a time reversal argument. Denote

$$\eta_n = \mathbf{1}(X_{n\Delta} < \alpha)\mathbf{1}(X_{(n+1)\Delta} > \alpha)(X_{(n+1)\Delta} - X_{n\Delta})^2.$$

We have

$$\mathbb{E}_{\sigma,b}\Big[\Big(\sum_{n=0}^{N-1} \mathbf{1}(X_{n\Delta} < \alpha)\mathbf{1}(X_{(n+1)\Delta} > \alpha)(X_{(n+1)\Delta} - X_{n\Delta})^2\Big)^2\Big] = \sum_{n=0}^{N-1}\mathbb{E}_{\sigma,b}[\eta_n^2] + 2\sum_{0\leq n<m}^{N-1}\mathbb{E}_{\sigma,b}[\eta_n\eta_m].$$

Denote by p_t the transition kernel of the diffusion X. Uniform bounds on diffusion coefficients imply that

$$p_t(x,y) \leq M_1\frac{1}{\sqrt{t}}\exp\Big(-\frac{(x-y)^2}{M_2 t}\Big), \tag{4.60}$$

with M_1, M_2 positive constants uniform on Θ, see Theorem 2.12 from Chapter 2, compare with [75, Lemma 2]. From (4.60) and the inequality [1, Formula 7.1.13]:

$$\int_x^\infty e^{-z^2}dz \leq \frac{e^{-x^2}}{x+\sqrt{x^2+4/\pi}} \leq \frac{\sqrt{\pi}}{2}e^{-x^2}, \tag{4.61}$$

follows that

$$\begin{aligned}
&\int_0^\alpha \int_\alpha^1 p_\Delta(x,y)(y-x)^4 dydx \\
&\quad \lesssim \int_0^\alpha \int_\alpha^1 \frac{1}{\sqrt{\Delta}} e^{-\frac{(y-x)^2}{c\Delta}}(y-x)^4 dydx \lesssim \Delta^2 \int_0^\alpha \int_{\frac{\alpha-x}{\sqrt{\Delta c}}}^{\frac{1-x}{\sqrt{\Delta c}}} e^{-z^2} z^4 dzdx \\
&\quad \lesssim \Delta^2 \int_0^\alpha \int_{\frac{\alpha-x}{\sqrt{\Delta c}}}^{\frac{1-x}{\sqrt{\Delta c}}} e^{-\frac{z^2}{2}} dzdx \lesssim \Delta^2 \int_0^\alpha e^{-\frac{(\alpha-x)^2}{2c\Delta}} dx \lesssim \Delta^{5/2}.
\end{aligned} \tag{4.62}$$

Similarly

$$\int_0^\alpha \int_\alpha^1 p_\Delta(x,y)(y-x)^2 dydx \lesssim \Delta^{3/2}. \tag{4.63}$$

For simplicity we will use the stationarity of X, which is granted by Assumption 4.2. Using more elaborated arguments the result could be obtained for an arbitrary initial condition. By stationarity, for any t, the one dimensional margin X_t is distributed with respect to the invariant measure $\mu(x)dx$. Conditioning on $X_{n\Delta}$, from (4.62) and uniform bounds on the density μ it follows

$$\mathbb{E}_{\sigma,b}[\eta_n^2] = \int_0^\alpha \int_\alpha^1 p_\Delta(x,y)(y-x)^4 dy\mu(x)dx \lesssim \Delta^{5/2}.$$

Hence

$$\sum_{n=0}^{N-1} \mathbb{E}_{\sigma,b}[\eta_n^2] \lesssim N\Delta^{\frac{5}{2}} = \Delta^{\frac{3}{2}}.$$

The Cauchy-Schwarz inequality implies

$$\sum_{n=0}^{N-2} \mathbb{E}_{\sigma,b}[\eta_n\eta_{n+1}] \lesssim \sum_{n=0}^{N-2} \mathbb{E}_{\sigma,b}[\eta_n^2]^{\frac{1}{2}} \mathbb{E}_{\sigma,b}[\eta_{n+1}^2]^{\frac{1}{2}} \lesssim N\Delta^{\frac{5}{2}} \lesssim \Delta^{\frac{3}{2}}.$$

Finally, using (4.63), for $m > n+1$, we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b}[\eta_n\eta_m] &= \int_0^\alpha \int_\alpha^1 \int_0^\alpha \int_\alpha^1 p_\Delta(x,y)(y-x)^2 p_{(m-n-1)\Delta}(z,x)(z-w)^2 p_\Delta(w,z)\mu(w)dydxdzdw \\
&\lesssim \int_0^\alpha \int_\alpha^1 p_\Delta(x,y)(y-x)^2 dydx \frac{1}{\sqrt{(m-n-1)\Delta}} \int_0^\alpha \int_\alpha^1 (z-w)^2 p_\Delta(w,z)dzdw \\
&\lesssim \Delta^{5/2} \frac{1}{\sqrt{m-n-1}}.
\end{aligned}$$

Consequently

$$\sum_{n=0}^{N-3}\sum_{m=n+2}^{N-1}\mathbb{E}_{\sigma,b}[\eta_n\eta_m]\lesssim\Delta^{5/2}\sum_{n=0}^{N-3}\sum_{k=1}^{N-n-2}\frac{1}{\sqrt{k}}\lesssim\Delta^{5/2}\sum_{n=0}^{N-3}\sqrt{N-n-2}$$
$$=\Delta^{5/2}\sum_{n=1}^{N-2}\sqrt{n}\lesssim\Delta^{5/2}N^{3/2}=\Delta. \qquad \square$$

Note that the claim of Theorem 4.41 still holds when we replace $(X_{(n+1)\Delta}-X_{n\Delta})^2$ by $(X_{(n+1)\Delta}-\alpha)^2$ or $(X_{n\Delta}-\alpha)^2$. Next, we show that, when considering the direction of the crossings, cancellations occur that make the difference of $\sum_{n=0}^{N-1}\chi(n,\alpha)(X_{(n+1)\Delta}-\alpha)^2$ and $\sum_{n=0}^{N-1}\chi(n,\alpha)(X_{n\Delta}-\alpha)^2$ even smaller.

Theorem 4.42. For any $\alpha\in[\frac{1}{J},1-\frac{1}{J}]$ we have

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1}\cdot\Big|\sum_{n=0}^{N-1}\chi(n,\alpha)\big((X_{(n+1)\Delta}-\alpha)^2-(X_{n\Delta}-\alpha)^2\big)\Big|^2\Big]^{\frac{1}{2}}\lesssim\Delta^{2/3}.$$

Due to the sign of the terms the proof of the next theorem cannot be done in a similar way as the previous result. In what follows we show that on the event $\mathcal{R}_1$

$$\sum_{n=0}^{N-1}\chi(n,\alpha)\big((X_{(n+1)\Delta}-\alpha)^2-(X_{n\Delta}-\alpha)^2\big)=\int_0^1\mathbf{1}(X_s<\alpha)ds-\frac{1}{N}\sum_{n=0}^{N-1}(\mathbf{1}(X_{n\Delta}<\alpha)+R,$$

where the remainder term is of the right order. Thus we are left with showing that

$$\mathbb{E}_{\sigma,b}\Big[\Big|\int_0^1\mathbf{1}(X_s<\alpha)ds-\frac{1}{N}\sum_{n=0}^{N-1}(\mathbf{1}(X_{n\Delta}<\alpha)\Big|^2\Big]^{\frac{1}{2}}\lesssim\Delta^{2/3}. \tag{4.64}$$

Note that $\frac{1}{N}\sum_{n=0}^{N-1}(\mathbf{1}(X_{n\Delta}<\alpha)$ is a Riemann type estimator of the occupation time of the interval $[0,\alpha)$. The problem of establishing the rate of convergence was recently considered in [64, 58]. Although obtained results do not apply as they require higher smoothness of the coefficients, they suggest an ever better rate $\Delta^{3/4}$. Indeed, in the case of reflected diffusion with bounded coefficients, we can show that

$$\mathbb{E}_{\sigma,b}\Big[\Big|\int_0^1 f(X_s)ds-\frac{1}{N}\sum_{n=0}^{N-1}f(X_{n\Delta})\Big|^2\Big]^{\frac{1}{2}}\lesssim\Delta^{\frac{1+s}{2}}\|f\|_{H^s},$$

for any function f with Sobolev regularity $0\le s\le 1$, see Corollary 5.8 in Chapter 5.

Proof. Fix $\alpha\in[\frac{1}{J},1-\frac{1}{J}]$. On the event $\mathcal{R}_1$, whenever $\mathbf{1}_{[0,\alpha)}(X_{(n+1)\Delta})-\mathbf{1}_{[0,\alpha)}(X_{n\Delta})\neq 0$ we must have $|X_{n\Delta}-\alpha|,|X_{(n+1)\Delta}-\alpha|\le\omega(\Delta)<\Delta^{4/9}$. Consider function $d:[0,1]\to\mathbb{R}$ given by

$$d(x)=(x-\alpha)^2\mathbf{1}(|x-\alpha|\le\Delta^{4/9})+\Delta^{8/9}\mathbf{1}(|x-\alpha|>\Delta^{4/9}).$$

We have

$$(\mathbf{1}_{[0,\alpha)}(X_{(n+1)\Delta})-\mathbf{1}_{[0,\alpha)}(X_{n\Delta}))\big((X_{(n+1)\Delta}-\alpha)^2-(X_{n\Delta}-\alpha)^2\big)=$$

$$= \big(\mathbf{1}_{[0,\alpha)}(X_{(n+1)\Delta}) - \mathbf{1}_{[0,\alpha)}(X_{n\Delta})\big)\big(d(X_{(n+1)\Delta}) - d(X_{n\Delta})\big).$$

Step 1. We will first show that

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \Big|\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta})\big(d(X_{(n+1)\Delta}) - d(X_{n\Delta})\big)\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{2/3}. \tag{4.65}$$

Note that

$$\begin{aligned} d'(x) &= 2(x-\alpha)\mathbf{1}(|x-\alpha| \le \Delta^{4/9}), \\ \frac{1}{2}d''(x) &= -\Delta^{4/9}\delta_{\{\alpha-\Delta^{4/9}\}} + \mathbf{1}(|x-\alpha| \le \Delta^{4/9}) - \Delta^{4/9}\delta_{\{\alpha+\Delta^{4/9}\}}, \end{aligned}$$

where the second derivative must be understood in the distributional sense. Since we fixed α separated from the boundaries, $d'(0) = d'(1) = 0$ for Δ small enough. Denote by

$$L_{s,t}(x) := L_t(x) - L_s(x),$$

the local time of the path fragment $(X_u, s \le u \le t)$. From the Itô-Tanaka formula [73, Chapter VI, Theorem 1.5] it follows that

$$\begin{aligned} d(X_{(n+1)\Delta}) - d(X_{n\Delta}) = \int_{n\Delta}^{(n+1)\Delta} d'(X_s)\sigma(X_s)dW_t + \int_{n\Delta}^{(n+1)\Delta} d'(X_s)b(X_s)ds + \\ + \int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)\mathbf{1}(|X_s-\alpha| \le \Delta^{4/9})ds - \Delta^{4/9}L_{n\Delta,(n+1)\Delta}(\alpha - \Delta^{4/9}) \\ - \Delta^{4/9}L_{n\Delta,(n+1)\Delta}(\alpha + \Delta^{4/9}) := \int_{n\Delta}^{(n+1)\Delta} d'(X_s)\sigma(X_s)dW_t + D_n. \end{aligned}$$

First, we will bound the sum of the martingale terms. Since martingale increments are uncorrelated, using Itô isometry, we obtain that

$$\begin{aligned} &\mathbb{E}_{\sigma,b}\Big[\Big|\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta}) \int_{n\Delta}^{(n+1)\Delta} d'(X_s)\sigma(X_s)dW_t\Big|^2\Big] = \\ &= \sum_{n=0}^{N-1} \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{[0,\alpha)}(X_{n\Delta}) \int_{n\Delta}^{(n+1)\Delta} (d'(X_s)\sigma(X_s))^2 ds\Big] \\ &\lesssim \Delta^{\frac{8}{9}}\mathbb{E}_{\sigma,b}\Big[\int_0^1 \mathbf{1}(|X_s-\alpha| \le \Delta^{\frac{4}{9}})ds\Big] = \Delta^{\frac{8}{9}} \int_{\alpha-\Delta^{\frac{4}{9}}}^{\alpha+\Delta^{\frac{4}{9}}} \mathbb{E}_{\sigma,b}[\mu_1(x)]dx \lesssim \Delta^{\frac{4}{3}}, \end{aligned}$$

where the last inequality follows from (4.15). Now, we will bound the sum of the finite variation terms: $\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta})D_n$. Note first, that since b is uniformly bounded, we have

$$\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta})\Big|\int_{n\Delta}^{(n+1)\Delta} d'(X_s)b(X_s)ds\Big| \lesssim \Delta^{4/9} \int_0^1 \mathbf{1}(|x-\alpha| \le \Delta^{4/9})\mu_1(x)dx \lesssim \Delta^{8/9}\|\mu_1\|_\infty.$$

Since by the inequality (4.15) $\|\mu_1\|_\infty$ has all moments finite, the root mean squared value of this sum is of smaller order than $\Delta^{2/3}$. Now, note that since on the event $\mathcal{R}_1$ $\omega(\Delta) <$

$\Delta^{4/9}$, condition $X_{n\Delta} < \alpha$ implies $L_{n\Delta,(n+1)\Delta}(\alpha + \Delta^{4/9}) = 0$. On the other hand, whenever $L_{n\Delta,(n+1)\Delta}(\alpha - \Delta^{4/9}) \neq 0$ we must have $X_{n\Delta} < \alpha$. Hence

$$\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta})(\Delta^{4/9} L_{n\Delta,(n+1)\Delta}(\alpha - \Delta^{4/9}) + \Delta^{4/9} L_{n\Delta,(n+1)\Delta}(\alpha + \Delta^{4/9})) = \Delta^{4/9} L_1(\alpha - \Delta^{4/9}).$$

Using first the Cauchy-Schwarz inequality and then the regularity of the local time (see [73, Chapter VI, Corollary 1.8 and the remark before]) we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b}\Big[\big|\Delta^{4/9} L_1(\alpha - \Delta^{4/9}) - \int_{\alpha-\Delta^{4/9}}^{\alpha} L_1(x)dx\big|^2\Big] \\
&\leq \Delta^{4/9} \int_{\alpha-\Delta^{4/9}}^{\alpha} \mathbb{E}_{\sigma,b}[|L_1(x) - L_1(\alpha - \Delta^{4/9})|^2]dx \\
&\lesssim \Delta^{4/9} \int_{\alpha-\Delta^{4/9}}^{\alpha} |x - (\alpha - \Delta^{4/9})|dx \lesssim \Delta^{4/3}.
\end{aligned}$$

Consequently, to prove (4.65) we just have to argue that the root mean squared error of

$$\begin{aligned}
\int_{\alpha-\Delta^{4/9}}^{\alpha} L_1(x)dx - \sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{n\Delta}) \int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)\mathbf{1}(|X_s - \alpha| \leq \Delta^{4/9})ds \\
= \sum_{n=0}^{N-1} \int_{n\Delta}^{(n+1)\Delta} (\mathbf{1}(X_s < \alpha) - \mathbf{1}(X_{n\Delta} < \alpha))\sigma^2(X_s)\mathbf{1}(|X_s - \alpha| \leq \Delta^{4/9})ds \\
= \sum_{n=0}^{N-1} \int_{n\Delta}^{(n+1)\Delta} (\mathbf{1}(X_s < \alpha) - \mathbf{1}(X_{n\Delta} < \alpha))\sigma^2(X_s)ds \qquad (4.66)
\end{aligned}$$

is of order $\Delta^{2/3}$. From the $1/2-$Hölder property of σ^2 it follows that

$$\begin{aligned}
\Big|\sum_{n=0}^{N-1} \int_{n\Delta}^{(n+1)\Delta} (\mathbf{1}(X_s < \alpha) - \mathbf{1}(X_{n\Delta} < \alpha))(\sigma^2(X_s) - \sigma^2(\alpha))ds\Big| \lesssim \\
\lesssim \sum_{n=0}^{N-1} \int_{n\Delta}^{(n+1)\Delta} \mathbf{1}(|X_s - \alpha| \leq \Delta^{\frac{4}{9}})\Delta^{\frac{2}{9}} ds \lesssim \Delta^{\frac{2}{9}} \int_{\alpha-\Delta^{4/9}}^{\alpha+\Delta^{4/9}} \mu_1(dx) \lesssim \Delta^{\frac{2}{3}} \|\mu_1\|_\infty.
\end{aligned}$$

Thus, by (4.15), we reduced (4.66) to

$$\int_0^1 \mathbf{1}(X_s < \alpha)ds - \frac{1}{N} \sum_{n=0}^{N-1} (\mathbf{1}(X_{n\Delta} < \alpha),$$

which is of the right order by (4.64). We conclude that (4.65) holds.

Step 2. Consider the time reversed process $Y_t = X_{1-t}$. Since X is reversible, the process Y, under the measure $\mathbb{P}_{\sigma,b}$, has the same law as X. Furthermore, the occupation density and the modulus of continuity of processes Y and X are identical, hence $\mathcal{R}_1$ is a "good" event also for Y. Inequality (4.65) is equivalent to

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{R}_1} \cdot \Big| \sum_{m=0}^{N-1} \mathbf{1}_{[0,\alpha)}(Y_{m\Delta})(d(Y_{(m+1)\Delta}) - d(Y_{m\Delta}))\Big|^2\Big]^{\frac{1}{2}} \lesssim \Delta^{\frac{2}{3}}.$$

Substituting $n = N - m$ we obtain

$$\sum_{m=0}^{N-1} \mathbf{1}_{[0,\alpha)}(Y_{m\Delta})(d(Y_{(m+1)\Delta}) - d(Y_{m\Delta})) = -\sum_{n=0}^{N-1} \mathbf{1}_{[0,\alpha)}(X_{(n+1)\Delta})(d(X_{(n+1)\Delta}) - d(X_{n\Delta})). \qquad \square$$

4.4. Low-frequency analysis

4.4.1. Spectral estimation method

In 1998 Hansen et al. [46] explained how the coefficients of a diffusion process are related to the spectral properties of its infinitesimal generator. In this section, we want to shortly introduce the main idea of their method.

The generator L of the reflected diffusion X is an unbounded operator on L^2 with

$$\begin{aligned} \operatorname{dom}(L) &= \{f \in H^2 : f'(0) = f'(1) = 0\} \\ Lf(x) &= \mu^{-1}(x)\big(\frac{1}{2}\sigma^2(x)\mu(x)f'(x)\big)', \text{ for f} \in \operatorname{dom}(T). \end{aligned}$$

Spectral properties of L are discussed in the Appendix A.2. Seen as an operator on the equivalent Hilbert space $L^2(\mu)$, the generator L is elliptic, self-adjoint and has a compact resolvent operator. Consequently, the eigenproblem

Eigenproblem 4.43. Find $(\zeta, u) \in \mathbb{R} \times L^2$, with $u \neq 0$, such that

$$Lu = \zeta u.$$

has countably many non-positive eigenvalues $0 = \zeta_0 > \zeta_1 > \zeta_2 \geq ...$, with $\mu-$orthogonal eigenfunctions $(u_i)_{i=0,...}$. The eigenvalue ζ_1 is simple and the corresponding eigenfunction u_1 is strictly monotone, see Proposition A.5. The main idea of the spectral estimation method is that the diffusion coefficient σ^2 can be expressed in terms of the invariant density μ and the eigenpair (ζ_1, u_1) (c.f. Hansen et al. [46, Eq. 5.2]):

$$\sigma^2(x) = \frac{2\zeta_1 \int_0^x u_1(y)\mu(y)dy}{u_1'(x)\mu(x)}. \tag{4.67}$$

4.4.2. Estimation error of the invariant measure

From now on we take the Assumptions 4.2 and 4.1 as granted. Fix $\Delta > 0$ and $0 < a < b < 1$. Set $J \sim N^{1/5}$. Since the generator L has a spectral gap, diffusion X is geometrically ergodic. Below, we state general bounds on the variance of integrals with respect to the empirical measure $\widehat{\mu}_N$, which are due to the mixing property of the observed sample $(X_{n\Delta})_{n=0,...,N}$. For the proof we refer to Chorowski and Trabs [23, Lemma 10].

Lemma 4.44. *For any $v, u \in L^2([0,1])$ we have*

$$\mathrm{Var}_{\sigma,b}\Big[\frac{1}{N}\sum_{n=0}^{N-1} v(X_{n\Delta})\Big] \lesssim N^{-1}\|v\|_{L^2}^2,$$

$$\mathrm{Var}_{\sigma,b}\Big[\frac{1}{N}\sum_{n=0}^{N-1} v(X_{n\Delta})u(X_{(n+1)\Delta})\Big] \lesssim N^{-1}\|v \cdot P_\Delta u\|_{L^2}^2.$$

Corollary 4.45. *There exists a high probability event $\mathcal{T}_1$, with $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_1) \lesssim N^{-1}J^2$, such that, for any $1 \le j \le J$, on $\mathcal{T}_1$ we have*

$$J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widehat{\mu}_N(dx) \sim 1.$$

Proof. Since the invariant density μ is uniformly bounded on Θ, there exist constants $0 < c < C$ such that $c \le J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \mu(x)dx \le C$. Let

$$\mathcal{T}_1 = \Big\{\forall j = 1, ..., J \text{ it holds } \Big|\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx) - \int_0^1 \psi_j'(x)\mu(x)dx\Big| \le \frac{c}{2J}\Big\}.$$

By Markov's inequality

$$\begin{aligned}\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_1) &\le \sum_{j=1}^J \mathbb{P}_{\sigma,b}\Big(\Big|\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx) - \int_0^1 \psi_j'(x)\mu(x)dx\Big| > \frac{c}{2J}\Big) \\ &\lesssim J^2\sum_{j=1}^J \mathrm{Var}_{\sigma,b}\Big(\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)\Big) \lesssim J^2N^{-1},\end{aligned}$$

where we used Lemma 4.44 with $\|\psi_j'\|_{L^2}^2 = J^{-1}$. □

4.4.3. Estimation error of the eigenpair (κ_1, u_1)

First, we state the approximation properties of spaces V_J.

Definition 4.46. Denote by π_J and π_J^μ the L^2 and $L^2(\mu)-$orthogonal projections on V_J respectively.

Since V_J is the space of linear spline functions with regular knots at $\{0, \frac{1}{J}, \frac{2}{J}, ..., \frac{J-1}{J}, 1\}$, it satisfies the following Jackson and Bernstein type inequalities:

$$\|(I-\pi_J)f\|_{H^k} \lesssim J^{-(2-k)\alpha}\|f\|_{C^{1,\alpha}} \text{ for } f \in C^{1,\alpha}([0,1]) \text{ and } k = 0, 1, \quad (4.68)$$

$$\|v\|_{H^1} \lesssim J\|v\|_{L^2} \text{ for } v \in V_J. \quad (4.69)$$

Definition 4.47. Denote by $(\phi_j)_{j=0,...,J}$ the Franklin system on $[0,1]$, i.e. the L^2-orthogonal basis of V_J, obtained from the Schauder algebraic basis by the Gram–Schmidt orthonormalization procedure.

For construction and properties of the Franklin system refer to [24]. In particular, basis functions $(\phi_j)_j$ satisfy the following uniform bound (cf. [24, Theorem 5]):

$$\Big\| \sum_{j=0}^{J} \phi_j^2 \Big\|_\infty \lesssim J. \tag{4.70}$$

Consider finite dimensional eigenvalue problem for the operator P_Δ on V_J:

Eigenproblem 4.48. Find $(\kappa_J, u_J) \in \mathbb{R} \times V_J$, with $u \neq 0$, such that

$$\langle P_\Delta u_J, v\rangle_\mu = \kappa_J \langle u_J, v\rangle_\mu \text{ for all } v \in V_J.$$

Remark 4.49. The eigenpairs of the self-adjoint operator $\pi_J^\mu P_\Delta \pi_J^\mu$ solve the Eigenproblem 4.48. In particular, 1 is the biggest eigenvalue corresponding to the constant eigenfunction.

Proposition 4.50. *Let $(\kappa_{J,i}, u_{J,i})_{i=0,\dots,J}$, with $1 = \kappa_{J,0} \geq \kappa_{J,1} \geq \dots \geq \kappa_{J,J}$ and $\|u_{J,i}\|_{L^2} = 1$, be solutions of the Eigenproblem 4.48. Then, for sufficiently large J, the following bound holds uniformly on Θ:*

$$|\kappa_1 - \kappa_{J,1}| + \|u_1 - u_{J,1}\|_{H^1} \lesssim J^{-1}. \tag{4.71}$$

Furthermore, the Eigenproblem 4.48 has a uniform spectral gap, i.e.

$$\min(1 - \kappa_{J,1}, \kappa_{J,1} - \kappa_{J,2}) \gtrsim 1.$$

Proof. The idea of the proof is the same as in [23, Proposition 12]. Denote by

$$L_0^2 = \Big\{ v \in L^2 : \int_0^1 v(x)\mu(x)dx = 0 \Big\}$$

the $\mu-$orthogonal complement of constant functions. By Remark 4.49 solutions $(\kappa_{J,i}, u_{J,i})_i$ are eigenpairs of the compact, self-adjoint and positive definite operator $\pi_J^\mu P_\Delta \pi_J^\mu$ on the Hilbert space $L_0^2(\mu)$, with the inner product $\langle \cdot, \cdot \rangle_\mu$. Since the transition operator P_Δ inherits the spectral properties of the generator L, it has a uniform spectral gap on Θ and $u_1 \in C^{1,1}([0,1])$ with $\|u_1\|_{C^{1,1}} \lesssim 1$ (see Proposition A.6). For J large enough, by Jackson's inequality (4.68), the assumptions of Theorem A.1 are fulfilled. It follows that

$$|\kappa_{J,1} - \kappa_1| + \Big\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \Big\|_{L^2(\mu)} \lesssim \|(I - \pi_J^\mu)u_1\|_{L^2(\mu)}. \tag{4.72}$$

By the equivalence of norms $\|\cdot\|_{L^2(\mu)}$ and $\|\cdot\|_{L^2}$ it holds

$$\|(I - \pi_J^\mu)u_1\|_{L^2(\mu)} = \|(I - \pi_J^\mu)(I - \pi_J)u_1\|_{L^2(\mu)} \lesssim \|(I - \pi_J)u_1\|_{L^2} \lesssim J^{-2}.$$

Since $\|u_1 - u_{J,1}\|_{L^2} \lesssim \Big\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \Big\|_{L^2(\mu)}$, from (4.72) we deduce that

$$|\kappa_{J,1} - \kappa_1| + \|u_{J,1} - u_1\|_{L^2} \lesssim J^{-2}.$$

Bernstein's (4.69) and Jackson's (4.68) inequalities imply that

$$\begin{aligned} \|u_1 - u_{J,1}\|_{H^1} &\leq \|(I - \pi_J)u_1\|_{H^1} + \|\pi_J u_1 - u_{J,1}\|_{H^1} \lesssim J^{-1} + J\,\|\pi_J u_1 - u_{J,1}\|_{L^2} \\ &\lesssim J^{-1} + J\,\|(I - \pi_J)u_1\|_{L^2} + J\,\|u_1 - u_{J,1}\|_{L^2} \lesssim J^{-1}, \end{aligned}$$

hence (4.71) holds.

Finally, we have to prove the uniform lower bound on the spectral gap. Theorem A.1 implies $\kappa_{J,1} - \kappa_{J,2} \gtrsim 1$. By the uniform spectral gap of the generator L holds $1 - \kappa_1 \gtrsim 1$. Furthermore, since $|\kappa_{J,1} - \kappa_1| \lesssim J^{-2}$, for J large enough we must have $1 - \kappa_{J,1} \gtrsim 1$. □

Proposition 4.51. *There exists a high probability event $\mathcal{T}_2$, with $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1}J^3$, such that*

$$\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{T}_2} \cdot (|\kappa_{J,1} - \widehat{\kappa}_{J,1}|^2 + \|u_{J,1} - \widehat{u}_{J,1}\|_{L^2}^2)\big] \lesssim N^{-1}J \tag{4.73}$$

holds uniformly on Θ.

Proof. Consider matrices $\widehat{P}_J = (\widehat{P}_{i,j})$ and $\widehat{G}_J = (\widehat{G}_{i,j})$ that correspond to forms $\widehat{p}$ and $\widehat{g}$ with respect to the basis $(\phi_j)_j$, i.e. $\widehat{P}_{i,j} = \widehat{p}(\phi_i, \phi_j)$ and $\widehat{G}_{i,j} = \widehat{g}(\phi_i, \phi_j)$. Let $P_J = (P_{i,j})$ and $G_J = (G_{i,j})$ be respectively: the matrix representation of the action of the transition operator P_Δ on $(\phi_j)_j$ and the Gram matrix of the $\mu-$induced inner product; i.e.

$$P_{i,j} = \langle P_\Delta \phi_i, \phi_j \rangle_\mu \quad \text{and} \quad G_{i,j} = \langle \phi_i, \phi_j \rangle_\mu .$$

Since the basis $(\phi_j)_j$ is an orthogonal system for $L^2([0,1])$, there exists an isometry between functions $v \in V_J$ and the coefficient vectors $(\langle v, \phi_j \rangle)_j$. To simplify the notation, we will throughout use v to denote both the function and the corresponding vector. Similarly, we will use $\|\cdot\|_{L^2}$ to denote both the $L^2([0,1])$ norm and the standard Euclidean norm on $\mathbb{R}^{J+1}$.

Using uniform variance bounds from Lemma 4.44 and the summation property (4.70) of the Franklin system $(\phi_j)_j$, we obtain that for any $v \in V_J$ (see [23, Lemmas 14 and 15])

$$\mathbb{E}_{\sigma,b}\big[\|(G_J - \widehat{G}_J)v\|_{L^2}^2 + \|(P_J - \widehat{P}_J)v\|_{L^2}^2\big] \lesssim N^{-1}J\|v\|_{L^2}^2. \tag{4.74}$$

Since the invariant density μ has a uniform positive lower bound on Θ, the matrix G_J is positive-definite with norm $\|G_J^{-1}\|_{L^2}$ uniformly bounded. By a standard Neumann series argument and Markov's inequality we obtain that, on a high probability event $\mathcal{T}_{2,1}$ (with $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_{2,1}) \lesssim N^{-1}J^2$)), matrix $\widehat{G}_J$ is invertible with $\|\widehat{G}_J^{-1}\|_{L^2} \lesssim 1$ (cf. [23, Lemma 6]). Form the residual vector

$$r = (\widehat{P}_J - P_J)u_{J,1} + \kappa_{J,1}(G_J - \widehat{G}_J)u_{J,1}.$$

Since by Remark 4.49 the eigenvalue $\kappa_{J,1} \leq 1$, from (4.74) it follows that

$$\mathbb{E}_{\sigma,b}[\|r\|_{L^2}^2] \lesssim N^{-1}J.$$

By Theorem A.10 there exists some $0 \leq i_0 \leq J$ such that the eigenpair $(\widehat{\kappa}_{J,i_0}, \widehat{u}_{J,i_0})$ satisfies

$$\begin{aligned} |\kappa_{J,1} - \widehat{\kappa}_{J,i_0}| &\leq \left\|\widehat{G}_J^{-1}\right\|_{L^2} \|r\|_{L^2} , \\ \|u_{J,1} - \widehat{u}_{J,i_0}\|_{L^2} &\leq \frac{2\sqrt{2}}{\delta(\widehat{\kappa}_{J,i})} \|\widehat{G}_J\|_{L^2}^{1/2} \|\widehat{G}_J^{-1}\|_{L^2}^{3/2} \|r\|_{L^2} , \end{aligned}$$

where $\delta(\widehat{\kappa}_{J,i_0}) = \min_{j \neq i_0}\{|\widehat{\kappa}_{J,j} - \kappa_{J,1}|\}$ is the so-called isolation distance of the eigenvalues $\widehat{\kappa}_{J,i_0}$ and $\kappa_{J,1}$. Let s be the uniform on Θ lower bound on the spectral gap of operators P_J (see Proposition 4.50). Define the event $\mathcal{T}_2$ as the subset of $\mathcal{T}_{2,1}$ for which $i_0 = 1$ and $\delta(\widehat{\kappa}_{J,1}) \geq s_1/2$. Since $\left\|\widehat{G}_J^{-1}\right\|_{L^2}$ and $\left\|\widehat{G}_J\right\|_{L^2}$ are uniformly bounded on the event $\mathcal{T}_{2,1}$, and since $\mathbb{E}_{\sigma,b}[\|r\|_{L^2}^2] \lesssim N^{-1}J$, inequality (4.73) holds.

To finish the proof we must show that $\mathcal{T}_2$ is a high probability event. Using Theorem A.11 we obtain that $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1}J^3$, exactly as in the proof of [23, Proposition 17]. □

4.4.4. Proof of Theorem 4.9

Choose $J \sim N^{1/5}$. Recall that the biggest negative eigenvalue of the generator is ζ_1, which is estimated by $\widehat{\zeta}_1 = \frac{\log(\widehat{\kappa}_1)}{\Delta}\mathbf{1}(0 < \widehat{\kappa}_1)$.

Corollary 4.52. *There exists a high probability event $\mathcal{T}_3 \subseteq \mathcal{T}_2$, with $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_3) \lesssim N^{-2/5}$, such that*

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_3} \cdot (|\zeta_1 - \widehat{\zeta}_1|^2 + \|u_1 - \widehat{u}_1\|_{H^1}^2)] \lesssim N^{-2/5}. \tag{4.75}$$

Furthermore, on the event $\mathcal{T}_3$, $|\widehat{\zeta}_1| \sim 1$ and $\|\widehat{u}_1\|_{H^1} \lesssim 1$.

Proof. Let $\mathcal{T}_2$ be the high probability event introduced in Proposition 4.51. From Propositions 4.50 and 4.51, using Bernstein's inequality, we obtain that

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_2} \cdot (|\kappa_1 - \widehat{\kappa}_1|^2 + \|u_1 - \widehat{u}_1\|_{H^1}^2)] \lesssim N^{-2/5}.$$

The uniform bounds on the eigenvalues of the generator carry over to the eigenvalues of the transition operator. Hence $c < \kappa_1 < 1 - c$ for some positive constant c. Let

$$\mathcal{T}_3 = \mathcal{T}_2 \cap \{|\kappa_1 - \widehat{\kappa}_1| \leq c/2\} \cap \{\|\widehat{u}_1\|_{H^1} \lesssim 2\|u_1\|_{H^1}\}.$$

Then, Markov's inequality implies $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_3) \lesssim N^{-2/5}$. Since the logarithm is uniformly Lipschitz on $[c/2, \infty)$, we conclude that $\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_3} \cdot |\zeta_1 - \widehat{\zeta}_1|^2] \lesssim N^{-2/5}$. □

Before we can prove Theorem 4.9, we need to face one more technical difficulty. Since the estimator $\widehat{u}_1$ converges to the eigenfunction u_1 in the sense of expected H^1 norm only, we cannot postulate a uniform positive lower bound on $\inf_{x\in[a,b]} \widehat{u}_1'(x)$. We show next, that this difficulty can be overcome by applying the threshold $\widehat{\sigma}_S^2 \wedge D$ (cf. [23, Lemma 20]).

Lemma 4.53. *There exists a high probability event $\mathcal{T}_4 \subset \mathcal{T}_3 \cap \mathcal{T}_1$, with $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_4) \lesssim N^{-2/5}$, such that on $\mathcal{T}_4$, for j s.t. $[\frac{j-1}{J}, \frac{j}{J}] \subset (a, b)$, we have*

$$\widehat{\sigma}_{S,j}^2 \wedge D = \frac{-2\widehat{v}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{(\widehat{u}_{1,j} \vee c_{a,b}) \int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)} \wedge D,$$

for a deterministic constant $c_{a,b} > 0$ satisfying $c_{a,b} \leq \inf_{x\in[a,b]} u_1'(x)$.

Proof. Fix j s.t. $[\frac{j-1}{J}, \frac{j}{J}] \subset (a, b)$. Note that by Definition 4.6

$$\widehat{\sigma}_{S,j}^2 \wedge D = \frac{-2\widehat{\zeta}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\left(\widehat{u}_{1,j} \vee \frac{-2\widehat{\zeta}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)D}\right) \int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)}.$$

We will show that, on some high probability event $\mathcal{T}_4$, we have

$$\frac{-2\widehat{\zeta}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)D} \geq c_{a,b},$$

for some uniform on Θ constant $c_{a,b} > 0$. By Corollaries 4.45 and 4.52 on $\mathcal{T}_1 \cap \mathcal{T}_3$ it holds

$$\frac{-2\widehat{\zeta}_1}{\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)D} \sim J.$$

Thus, we just have to find a lower bound on $J\int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)$. Note, that from spectral representation (4.67) and the uniform lower bound on $\inf_{x\in[a,b]} u_1'(x)$ it follows that

$$J\int_0^1 \psi_j(x)u_1(x)\mu(x)dx = \frac{J}{-2v_1}\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x)u_1'(x)\mu(x)dx \geq c$$

for some positive constant $c = c(a,b) > 0$. Let

$$\mathcal{T}_4 = \mathcal{T}_1 \cap \mathcal{T}_3 \cap \Big\{\Big|\int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx) - \int_0^1 \psi_j(x)u_1(x)\mu(x)dx\Big| \leq \frac{c}{2J}\Big\}.$$

Using Corollary 4.52 together with Lemma 4.44 we obtain that

$$\begin{aligned}
&\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_3}\cdot J\Big|\int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx) - \int_0^1 \psi_j(x)u_1(x)\mu(x)dx\Big|^2\Big] \\
&= \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_3}\cdot\Big|J\int_{\frac{j-1}{J}}^{\frac{j}{J}}\Big(\int_0^x \widehat{u}_1(y)\widehat{\mu}_N(dy) - \int_0^x u_1(y)\mu(y)dy\Big)dx\Big|^2\Big] \\
&\leq \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_3}\cdot J\int_{\frac{j-1}{J}}^{\frac{j}{J}}\Big|\int_0^x \widehat{u}_1(y)\widehat{\mu}_N(dy) - \int_0^x u_1(y)\mu(y)dy\Big|^2 dx\Big] \\
&\lesssim J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_3}\Big(\int_0^x |\widehat{u}_1(y) - u_1(y)|\widehat{\mu}_N(dy)\Big)^2 \\
&\qquad + \mathbf{1}_{\mathcal{T}_3}\Big(\int_0^x u_1(y)\widehat{\mu}_N(dy) - \int_0^x u_1(y)\mu(y)dy\Big)^2\Big]dx \\
&\lesssim \mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_3}\cdot\|\widehat{u}_1 - u_1\|_{H^1}^2] + J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \mathrm{Var}_{\sigma,b}\Big[\int_0^x u_1(y)\widehat{\mu}_N(dy)\Big]dx \lesssim N^{-2/5}.
\end{aligned}$$

By Markov's inequality, we conclude that $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_4) \lesssim N^{-2/5}$. □

Proof of Theorem 4.9. Denote

$$\widetilde{u}_{1,j}' = \widehat{u}_{1,j}' \vee c_{a,b}$$

and

$$\widetilde{\sigma}_j^2 = \frac{-2\widehat{\zeta}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\widetilde{u}_{1,j}'\int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)}, \quad \widetilde{\sigma}^2 = \sum_{j=1}^{J-1}\widetilde{\sigma}_j^2\mathbf{1}_j.$$

By Lemma 4.53, the uniform bound D on σ^2, and since $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{T}_4) \lesssim N^{-2/5}$, we just have to verify that

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_4}\cdot\|\sigma^2 - \widetilde{\sigma}^2\|_{L^2([a,b])}^2] \lesssim N^{-2/5}. \tag{4.76}$$

We will show below that, for every j s.t. $[\frac{j-1}{J}, \frac{j}{J}] \subset (a,b)$, it holds

$$\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_4} \cdot |\sigma_j^2 - \widetilde{\sigma}_j^2|^2] \lesssim \mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_4} \cdot J \int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{u}_{1,j}' - u_1'(x)|^2 dx\Big] + N^{-2/5}, \tag{4.77}$$

with

$$\sigma_j^2 = \frac{\int_{\frac{j-1}{J}}^{\frac{j}{J}} \sigma^2(x) u_1'(x)\mu(x)dx}{\int_{\frac{j-1}{J}}^{\frac{j}{J}} u_1'(x)\mu(x)dx} = \frac{-2\zeta_1 \int_0^1 \psi_j(x) u_1(x)\mu(x)dx}{\int_0^1 \psi_j'(x) u_1'(x)\mu(x)dx}.$$

Integrating both sides of (4.77) on $[\frac{j-1}{J}, \frac{j}{J}]$ and summing over all j with $[\frac{j-1}{J}, \frac{j}{J}] \subset (a,b)$ yields

$$\mathbb{E}_{\sigma,b}\Big[\mathbf{1}_{\mathcal{T}_4} \cdot \|\sigma_J^2 - \widetilde{\sigma}\|^2_{L^2([a,b])}\Big] \lesssim \mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_4} \cdot \|\widehat{u}_1' - u_1'\|^2_{L^2}] + N^{-2/5} \lesssim N^{-2/5},$$

with $\sigma_J^2(x) = \sum_{j=1}^J \sigma_j^2 \mathbf{1}_j(x)$. Since the Sobolev regularity of σ^2 yields $\|\sigma^2 - \sigma_J^2\|_\infty \lesssim N^{-1/5}$, the inequality (4.76) follows.

Proof of (4.77). Set j s.t. $[\frac{j-1}{J}, \frac{j}{J}] \subset (a,b)$. The uniform lower bounds on the derivative u_1' and on the invariant density μ implies

$$\int_0^1 \psi_j'(x) u_1'(x)\mu(x)dx \gtrsim J^{-1}.$$

Consequently,

$$\begin{aligned}
\Big|\sigma_j^2 - \widetilde{\sigma}_j^2\Big| &= \Big|\frac{2v_1 \int_0^1 \psi_j(x)u_1(x)\mu(x)dx}{\int_0^1 \psi_j'(x)u_1'(x)\mu(x)dx} - \frac{2\widehat{v}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\int_0^1 \psi_j'(x)\widetilde{u}_{1,j}'\widehat{\mu}_N(dx)}\Big| \\
&\lesssim J\Big|v_1 \int_0^1 \psi_j(x)u_1(x)\mu(x)dx - \widehat{v}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)\Big| \\
&\quad + J|\widetilde{\sigma}_j^2|\Big|\int_0^1 \psi_j'(x)\widetilde{u}_{1,j}'\widehat{\mu}_N(dx) - \int_0^1 \psi_j'(x)u_1'(x)\mu(x)dx\Big| \\
&:= A_{j,1} + A_{j,2}.
\end{aligned}$$

Note that, since $J\int_0^1 \psi_j(x)u_1(x)\mu(x)dx = J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \big(\int_0^x u_1(y)\mu(y)dy\big)dx \lesssim 1$, and since on the event $\mathcal{T}_3$ the estimator $\widehat{v}_1$ is uniformly bounded, we have

$$A_{j,1} \quad \lesssim \quad |v_1 - \widehat{v}_1| + J\Big|\int_0^1 \psi_j(x)u_1(x)\mu(x)dx - \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)\Big|.$$

Arguing as in Lemma 4.53, we obtain that $\mathbb{E}_{\sigma,b}[\mathbf{1}_{\mathcal{T}_4} \cdot A_{j,1}^2] \lesssim N^{-2/5}$. We now need to bound the second term $A_{j,2}$. On the event $\mathcal{T}_4$, by Corollary 4.45, holds

$$\begin{aligned}
\Big|\frac{-2\widehat{v}_1 \int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)}{\widetilde{u}_{1,j}' \int_0^1 \psi_j'(x)\widehat{\mu}_N(dx)}\Big| &\lesssim J\Big|\int_0^1 \psi_j(x)\widehat{u}_1(x)\widehat{\mu}_N(dx)\Big| \\
&= J\int_{\frac{j-1}{J}}^{\frac{j-1}{J}} \Big|\int_0^x \widehat{u}_1(y)\widehat{\mu}_N(dy)\Big|dx
\end{aligned}$$

$$\lesssim \|\widehat{u}_1\|_{H^1}.$$

Hence, on $\mathcal{T}_4$ the estimator $\widetilde{\sigma}_j^2$ is uniformly bounded. Consequently,

$$\begin{aligned} A_{j,2} &\lesssim J\Big|\int_{\frac{j-1}{J}}^{\frac{j}{J}} \widetilde{u}'_{1,j}\widehat{\mu}_N(dx) - \int_{\frac{j-1}{J}}^{\frac{j}{J}} u'_1(x)\mu(x)dx\Big| \\ &\lesssim J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widetilde{u}'_{1,j} - u'_1(x)|\widehat{\mu}_N(dx) + \Big|J\int_{\frac{j-1}{J}}^{\frac{j}{J}} u'_1(x)\widehat{\mu}_N(dx) - J\int_{\frac{j-1}{J}}^{\frac{j}{J}} u'_1(x)\mu(x)dx\Big|. \end{aligned}$$

By Lemma 4.44

$$\mathbb{E}_{\sigma,b}\Big[\Big|J\int_{\frac{j-1}{J}}^{\frac{j}{J}} u'_1(x)\widehat{\mu}_N(dx) - J\int_{\frac{j-1}{J}}^{\frac{j}{J}} u'_1(x)\mu(x)dx\Big|^2\Big] \lesssim N^{-1}J\|u'_1\|_\infty^2 \lesssim N^{-4/5}.$$

Set $x_j \in [\frac{j-1}{J}, \frac{j}{J}]$. Since u'_1 is uniformly Lipschitz, by Corollary 4.45, on the event $\mathcal{T}_4$, holds

$$\begin{aligned} J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widetilde{u}'_{1,j} - u'_1(x)|\widehat{\mu}_N(dx) &\leq J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widetilde{u}'_{1,j} - u'_1(x_j)|\widehat{\mu}_N(dx) + J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |u'_1(x_j) - u'_1(x)|\widehat{\mu}_N(dx) \\ &\lesssim |\widetilde{u}'_{1,j} - u'_1(x_j)| + J^{-1} \lesssim J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{u}'_{1,j} - u'_1(x)|dx + N^{-1/5}. \end{aligned}$$

Finally, by the Cauchy-Schwarz inequality

$$\mathbb{E}_{\sigma,b}\Big[\mathcal{T}_4 \cdot \Big(J\int_{\frac{j-1}{J}}^{\frac{j}{J}} |\widehat{u}'_{1,j} - u'_1(x)|dx\Big)^2\Big] \lesssim J\int_{\frac{j-1}{J}}^{\frac{j}{J}} \mathbb{E}_{\sigma,b}[|\widehat{u}'_{1,j} - u'_1(x)|^2]dx,$$

hence (4.77) holds. □

5. Estimating the occupation time

The problem of estimating the occupation time functional $\Gamma_T(f) = \int_0^T f(X_r)dr$ of a stationary time-reversible Markov process X is considered. Mean L^2 convergence rates that depend on the action of the infinitesimal generator on f are obtained. When X is a stationary diffusion and f has Sobolev regularity of order s, a convergence rate $n^{-(1+s)/2}$ is derived.

5.1. Introduction

For a $d-$dimensional càdlàg process $(X_t, 0 \leq t \leq T)$ we consider the problem of estimating the integral functional Γ_t, defined for $f : \mathbb{R}^d \to \mathbb{R}$ by

$$\Gamma_t\left(f\right) = \int_0^t f(X_r)dr, \text{ for } 0 \leq t \leq T, \tag{5.1}$$

whenever the above integral is well defined. Functional of the above type appear in many stochastic formulas and applications. If f is the indicator function of a domain D, then

$$\Gamma_t(f) = \int_0^t \mathbf{1}(X_r \in D)dr$$

corresponds to the chronological occupation time of D, which is of particular interest in many fields such as queueing theory, biology or finance.

When X is observed at discrete equidistant times k/n, with $k = 0, \ldots, \lfloor nT \rfloor$, it is natural to approximate Γ_t by a Riemann type estimator:

$$\widehat{\Gamma}_{t,n}(f) = \frac{1}{n}\sum_{k=1}^{\lfloor nt \rfloor} f(X_{\frac{k-1}{n}}). \tag{5.2}$$

As explained in Kohatsu-Higa et al. [58], when the function f is not assumed smooth the functional Γ_t is not regular. In such a case, the estimation error of Γ_t by $\widehat{\Gamma}_t$ can't be analyzed with classical techniques. In this chapter, under the assumption of X being a stationary time-reversible Markov process, we obtain mean L^2 convergence rates that depend on the action of the infinitesimal generator on f. More precisely, we prove that for some constant C_T it holds

$$\mathbb{E}\left[\left|\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)\right|^2\right]^{\frac{1}{2}} \leq C_T n^{-\frac{1+s}{2}} \|(I-L)^{s/2} f\|_{L^2(\mu)},$$

where L is the infinitesimal generator of X.

Until recently, not much was known about the estimation error of Γ_t. First results were probably obtained by Ngo and Ogawa [64], in the setting of one dimensional diffusion processes. The most important finding was the identification of the unintuitive rate $n^{-3/4}$ for

$f = \mathbf{1}_{[0,\infty)}$. This result raised the question how exactly the estimation error depends on the regularity of the function f. A partial answer was soon given in Kohatsu-Higa et al. [58]. By the means of Malliavin calculus, the rate $n^{-3/4}$ for the indicator function was confirmed to hold for diffusion processes with smooth coefficients. Furthermore, the authors established the relation between the convergence rate and the Hölder regularity of f.

A different approach was applied in the recent works by Ganychenko, Knopova and Kulik [2014, 2015, 2015]. The authors assumed that X is a Markov process with regular transition probabilities. Applying a modification of Dynkin's theory of continuous additive functionals (see Dynkin [31]), they reached the Hölder rates from Kohatsu-Higa et al. [58]. Furthermore, the authors specified conditions under which the rate $n^{-1/2}$ holds for twice integrable functions. Nevertheless, the rate $n^{-3/4}$ for the indicator function was not obtained.

5.2. Preliminaries

Let $(X_t, 0 \le t \le T)$ be a stationary Markov process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a Polish space E, equipped with its Borel $\sigma-$field $\mathcal{B}(E)$. Denote by μ the stationary measure and let $\langle \cdot, \cdot \rangle_\mu$ and $\| \cdot \|_{L^2(\mu)}$ be the $\mu-$induced inner product and norm on the space $L^2(E, \mu)$, respectively. For $f \in L^2(E, \mu)$ and $t \ge 0$ let

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$$

be the transition operator. The family $(P_t)_{t\ge 0}$ forms a strongly continuous semigroup of contractions. The semigroup $(P_t)_{t\ge 0}$ is related to an operator L, called the infinitesimal generator, defined by:

$$\begin{aligned} \mathrm{dom}(L) &= \left\{ f \in L^2(E, \mu) \, : \, \lim_{t\to 0} \frac{(P_t - I)f}{t} \quad \text{exists} \right\}, \\ Lf &= \lim_{t\to 0} \frac{(P_t - I)f}{t}, \quad \text{for } f \in \mathrm{dom}(L). \end{aligned} \tag{5.3}$$

Assumption 5.1. *We assume that X is a stationary Markov process with the infinitesimal generator L being a non-positive, self-adjoint operator on the Hilbert space $L^2(E, \mu)$.*

In the following, an important role will be played by the spectral theory of the infinitesimal generator L. Before we proceed, let us define the basic objects. For more details we refer the reader to D. Bakry and Ledoux [26, Section A.4].

Definition 5.2. Consider an increasing family $(H_\lambda)_{\lambda \ge 0}$ of closed linear subspaces of the Hilbert space $L^2(E, \mu)$, which is right-continuous in the sense that $\bigcap_{\lambda' > \lambda} H_{\lambda'} = H_\lambda$. Furthermore, we require that $\bigcup_{\lambda \ge 0} H_\lambda$ is dense in $L^2(E, \mu)$. The *spectral measure* is the family $(E_\lambda)_{\lambda \ge 0}$ of orthogonal projections $E_\lambda : L^2(E, \mu) \to H_\lambda$.

For any $f, g \in L^2(E, \mu)$ the map $\lambda \to \langle E_\lambda f, g \rangle_\mu$ is right-continuous and of bounded variation. Consequently, for any measurable function $\psi : [0, \infty) \to \mathbb{R}$ we may define the Stieltjes integral

$$\int_0^\infty \psi(\lambda) d\langle E_\lambda f, g \rangle_\mu,$$

which by duality defines a symmetric linear operator denoted by

$$\int_0^\infty \psi(\lambda) dE_\lambda = \Psi,$$

with domain

$$\operatorname{dom}(\Psi) = \left\{ f \in L^2(E,\mu) \, : \, \int_0^\infty \psi(\lambda)^2 d\langle E_\lambda f, f\rangle_\mu < \infty \right\}.$$

Moreover, the operator norm (possibly infinite) of Ψ is given by

$$\|\Psi\| = \sup_{f\in\operatorname{dom}(\Psi)} \frac{\|\Psi f\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} = \sup_{f\in\operatorname{dom}(\Psi)} \frac{\int_0^\infty \psi(\lambda)^2 d\langle E_\lambda f, f\rangle_\mu}{\int_0^\infty d\langle E_\lambda f, f\rangle_\mu}.$$

Under Assumption 5.1, the spectral decomposition theorem (see for example D. Bakry and Ledoux [26, Theorem A.4.2]) asserts that there exists a spectral measure $(E_\lambda)_{\lambda\geq 0}$ such that

$$-L = \int_0^\infty \lambda dE_\lambda.$$

The existence of the spectral measure $(E_\lambda)_{\lambda\geq 0}$ allows us to define fractional powers of the infinitesimal generator L.

Definition 5.3. For $s > 0$ let

$$(-L)^{s/2} = \int_0^\infty \lambda^{s/2} dE_\lambda,$$

with

$$\operatorname{dom}\left((-L)^{s/2}\right) = \left\{ f \in L^2(E,\mu) \, : \, \int_0^\infty \lambda^s d\langle E_\lambda f, f\rangle_\mu < \infty \right\}.$$

5.3. Convergence rates

Theorem 5.4. Grant Assumption 5.1 and let $\Gamma_T, \widehat{\Gamma}_T$ be defined as in (5.1) and (5.2). There exists a positive constant $C < \infty$ that does not depend on the time horizon T nor on the process X such that for all $0 \leq s \leq 1$ and $f \in \operatorname{dom}\left((I-L)^{s/2}\right)$ it holds

$$\mathbb{E}\left[\left|\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)\right|^2\right]^{\frac{1}{2}} \leq C\sqrt{1\vee T} n^{-\frac{1+s}{2}} \|(I-L)^{s/2} f\|_{L^2(\mu)}.$$

Proof. Fix $T < \infty$, $0 \leq s \leq 1$ and $f \in \operatorname{dom}\left((I-L)^{s/2}\right)$. Note first that

$$\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f) = \sum_{k=1}^{\lfloor nT\rfloor} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \left(f(X_r) - f(X_{\frac{k-1}{n}})\right) dr + \int_{\frac{\lfloor nT\rfloor}{n}}^{T} f(X_r) dr, \tag{5.4}$$

where the second term is of negligible order. Indeed, using the Cauchy-Schwarz inequality we obtain

$$\mathbb{E}\left[\left|\int_{\frac{\lfloor nT \rfloor}{n}}^{T} f(X_r)dr\right|^2\right] \leq (T - \tfrac{\lfloor nT \rfloor}{n}) \int_{\frac{\lfloor nT \rfloor}{n}}^{T} \mathbb{E}[f^2(X_r)]dr \leq n^{-2}\|f\|_{L^2}^2$$
$$= n^{-2}\int_0^\infty d\langle E_\lambda f, f\rangle_\mu \leq n^{-2}\int_0^\infty (1+\lambda)^s d\langle E_\lambda f, f\rangle_\mu$$
$$= n^{-2}\|(I-L)^{s/2}f\|_{L^2(\mu)}^2.$$

Next, we will bound the mean squared error of the sum in (5.4). Expanding the square we get

$$\mathbb{E}\Big[\Big|\sum_{k=1}^{\lfloor nT \rfloor}\int_{\frac{k-1}{n}}^{\frac{k}{n}} (f(X_r) - f(X_{\frac{k-1}{n}}))dr\Big|^2\Big] =$$
$$= \sum_{k,l=1}^{\lfloor nT \rfloor}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{l-1}{n}}^{\frac{l}{n}} \mathbb{E}\big[(f(X_h) - f(X_{\frac{k-1}{n}}))(f(X_r) - f(X_{\frac{l-1}{n}})\big]drdh. \tag{5.5}$$

We will first bound the sum of the diagonal terms, then show that the off-diagonal terms are non-positive, hence the overall error is dominated by the diagonal terms. Calculating the conditional expectations, we obtain that when $\frac{k}{n} \geq h > r > \frac{k-1}{n}$ it holds

$$\mathbb{E}\big[(f(X_h) - f(X_{\frac{k-1}{n}}))(f(X_r) - f(X_{\frac{k-1}{n}})\big]$$
$$= \mathbb{E}\big[(P_{h-r}f(X_r) - f(X_{\frac{k-1}{n}}))(f(X_r) - f(X_{\frac{k-1}{n}})\big]$$
$$= \langle P_{h-r}f, f\rangle_\mu - \langle P_{h-\frac{k-1}{n}}f, f\rangle_\mu - \langle f, P_{r-\frac{k-1}{n}}f\rangle_\mu + \langle f, f\rangle_\mu$$
$$= \langle P_{h-r}f - P_{h-\frac{k-1}{n}}f - P_{r-\frac{k-1}{n}}f + f, f\rangle_\mu$$
$$= \langle (P_{h-r} - I)f + (I - P_{h-\frac{k-1}{n}})f + (I - P_{r-\frac{k-1}{n}})f, f\rangle_\mu.$$

Consequently,

$$\sum_{k=1}^{\lfloor nT \rfloor}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{k-1}{n}}^{\frac{k}{n}} \mathbb{E}\big[(f(X_h) - f(X_{\frac{k-1}{n}}))(f(X_r) - f(X_{\frac{k-1}{n}}))\big]drdh =$$
$$= 2\sum_{k=1}^{\lfloor nT \rfloor}\Big\langle\Big(\int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{k-1}{n}}^{h}(P_{h-r} - I)drdh + \int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{k-1}{n}}^{h}(I - P_{h-\frac{k-1}{n}})drdh +$$
$$+ \int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{k-1}{n}}^{h}(I - P_{r-\frac{k-1}{n}})drdh\Big)f, f\Big\rangle_\mu$$
$$= 2\sum_{k=1}^{\lfloor nT \rfloor}\Big\langle\Big(\int_0^{\frac{1}{n}}\int_0^{h}(P_{h-r} - I)drdh + \int_0^{\frac{1}{n}} h(I - P_h)drdh + \int_0^{\frac{1}{n}} (\tfrac{1}{n} - r)(I - P_r)dr\Big)f, f\Big\rangle_\mu$$
$$= 2\lfloor nT \rfloor\Big\langle\varphi(-L)f, f\Big\rangle_\mu = 2\lfloor nT \rfloor\int_0^\infty \varphi(\lambda)d\langle E_\lambda f, f\rangle_\mu,$$

where

$$\varphi(\lambda) = \Big(\int_0^{\frac{1}{n}}\int_0^h (e^{-\lambda(h-r)} - 1)drdh + \int_0^{\frac{1}{n}} h(1-e^{-\lambda h})dh + \int_0^{\frac{1}{n}} (\tfrac{1}{n} - r)(1-e^{-\lambda r})dr\Big).$$

Since for any $0 \le s \le 1$ and $x > 0$ it holds $1 - e^{-x} \le x^s$, we have

$$|\varphi(\lambda)| \le 3n^{-(2+s)}\lambda^s.$$

We conclude that the sum of diagonal terms is bounded by

$$2\lfloor nT \rfloor \int_0^\infty |\varphi(\lambda)| d\langle E_\lambda f, f\rangle_\mu \le 6Tn^{-(1+s)}\int_0^\infty \lambda^s d\langle E_\lambda f, f\rangle_\mu = 6Tn^{-(1+s)}\|(-L)^{s/2}f\|_{L^2(\mu)}.$$

To finish the proof, we have to show that the off diagonal terms in (5.5) are non-positive. By calculating the conditional expectations, we obtain for $h > \frac{k-1}{n} \ge \frac{l}{n} > r > \frac{l-1}{n}$ that

$$\begin{aligned}
&\mathbb{E}\big[(f(X_h) - f(X_{\frac{k-1}{n}}))(f(X_r) - f(X_{\frac{l-1}{n}})\big] \\
&\quad = \mathbb{E}\big[(P_{h-\frac{k-1}{n}} - I)f(X_{\frac{k-1}{n}})(f(X_r) - f(X_{\frac{l-1}{n}}))\big] \\
&\quad = \mathbb{E}\big[P_{\frac{k-1}{n}-r}(P_{h-\frac{k-1}{n}} - I)f(X_r)(f(X_r) - f(X_{\frac{l-1}{n}}))\big] \\
&\quad = \langle P_{\frac{k-1}{n}-r}(P_{h-\frac{k-1}{n}} - I)f, f\rangle_\mu - \langle P_{r-\frac{l-1}{n}}P_{\frac{k-1}{n}-r}(P_{h-\frac{k-1}{n}} - I)f, f\rangle_\mu \\
&\quad = \langle P_{\frac{k-1}{n}-r}(P_{h-\frac{k-1}{n}} - I)f, (I - P_{r-\frac{l-1}{n}})f\rangle_\mu
\end{aligned}$$

where we used stationarity and the self-adjoint property of the semigroup operator. Since $P_{\frac{k-1}{n}-r}$ and $I - P_{r-\frac{l-1}{n}}$ are positive definite and $(P_{h-\frac{k-1}{n}} - I)$ is negative definite we conclude that the above inner product is non-positive. □

Definition 5.5. For $s > 0$ we define the Bessel potential spaces associated with the generator L by

$$\mathcal{D}^s = \operatorname{dom}\Big((I-L)^{s/2}\Big) = \Big\{f \in L^2(E,\mu) \,:\, \int_0^\infty (1+\lambda)^s d\langle E_\lambda f, f\rangle_\mu < \infty\Big\},$$

with the norm

$$\|f\|_{\mathcal{D}^s} = \Big\|(I-L)^{s/2} f\Big\|_{L^2(\mu)} = \Big(\int_0^\infty (1+\lambda)^s d\langle E_\lambda f, f\rangle_\mu\Big)^{1/2}.$$

As the fractional operators $(I-L)^{s/2}$ are closed in the graph norm, $(\mathcal{D}^s, \|\cdot\|_{\mathcal{D}^s})$ are real Hilbert spaces.

The definition of spaces $\mathcal{D}^s$ is natural as it is a straightforward generalization of the Bessel potential spaces for the Laplace operator. For any $f \in \operatorname{dom}((-L)^{s/2})$ it holds $f \in \mathcal{D}^s$ and $\Big\|(-L)^{s/2}f\Big\|_{L^2(\mu)} \le \|f\|_{D^s}$. Consequently, Theorem 5.4 states that for any $0 \le s \le 1$ the error $\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)$ is a bounded linear operator from space $\mathcal{D}^s$ into $L^2(\Omega, \mathbb{P})$ with norm bounded by $C_T n^{-(1+s)/2}$.

5.4. Examples

5.4.1. Ornstein-Uhlenbeck process

Let $(X_t, t \geq 0)$ be a stationary $d-$dimensional Ornstein-Uhlenbeck process, defined by the stochastic differential equation:

$$dX_t = -X_t dt + \sqrt{2} dW_t,$$

(where W_t is a standard $d-$dimensional Brownian motion) and with initial condition:

$$X_0 \stackrel{d}{=} \mathcal{N}(0, I).$$

The invariant measure μ of X is the standard $d-$dimensional normal distribution. The infinitesimal generator L of X can be well described by its spectral decomposition. Let

$$H_k(x) = \frac{1}{\sqrt{k!}} \int_{\mathbb{R}} (x + iy)^k d\mu(y), \ x \in \mathbb{R},$$

be the one dimensional Hermite polynomials. $d-$dimensional tensor products

$$H_{\mathbf{k}}(x) = \prod_{k=1}^{d} H_{k_i}(x_i), \ \mathbf{k} = (k_1, ..., k_d) \in \mathbb{N}^d, \ x = (x_1, ..., x_d) \in \mathbb{R}^d$$

are the eigenfunctions of L with eigenvalues $-\overline{\mathbf{k}} = -\sum_{k=1}^{d} k_i$. The infinitesimal generator L acts on $f \in L^2(\mathbb{R}^d, \mu)$ by

$$Lf = -\sum_{\mathbf{k} \in \mathbb{N}^d} \langle f, H_{\mathbf{k}} \rangle_\mu \overline{\mathbf{k}} H_{\mathbf{k}}$$

with

$$\operatorname{dom}(L) = \left\{ f \in L^2(\mathbb{R}^d, \mu) \ : \ \sum_{\mathbf{k} \in \mathbb{N}^d} \langle f, H_{\mathbf{k}} \rangle_\mu^2 \overline{\mathbf{k}}^2 < \infty \right\},$$

see D. Bakry and Ledoux [26, Chapter 2.7.1]. As such, X satisfies Assumption 5.1. Consequently, for $f \in \operatorname{dom}\left((-L)^{s/2}\right)$ it holds

$$\mathbb{E}\left[\left|\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)\right|^2\right]^{\frac{1}{2}} \leq C_T n^{-\frac{1+s}{2}} \|(-L)^{s/2} f\|_{L^2(\mu)} \leq C_T n^{-\frac{1+s}{2}} \|f\|_{\mathcal{D}^s}. \tag{5.6}$$

The Bessel potential spaces $\mathcal{D}^s$ that correspond to the Ornstein-Uhlenbeck process play an important role in Malliavin calculus (for details the reader is referred to Watanabe [94, Chapter 1.3], where they are called Sobolev fractional spaces). For $s = 1$ the space $\mathcal{D}^1$ can be described in terms of weak derivatives. Using integration by parts, one can show that (see e.g. Pavliotis [67, Proposition 4.1 Eq. (4.45)]) for $f \in C_c^2(\mathbb{R}^d)$ it holds

$$\|f\|_{\mathcal{D}^1}^2 = \langle f, (I - L) f \rangle_\mu = \|f\|_{L^2(\mu)}^2 + \frac{1}{2} \|\nabla f\|_{L^2(\mu)}^2. \tag{5.7}$$

Since $C_c^2(\mathbb{R}^d)$ is dense in $\mathcal{D}^1$, we can identify $\mathcal{D}^1$ with the space of weakly differentiable functions f such that $f, f' \in L^2(\mathbb{R}^d, \mu)$.

In the following, we will deduce from inequality (5.6) similar upper error bounds for standard Sobolev spaces on $\mathbb{R}^d$.

Definition 5.6. For $s > 0$ we define the Sobolev norm of order s by

$$\|f\|_{H^s} = \frac{1}{(2\pi)^{d/2}} \Big(\int_{\mathbb{R}^d} (1+|u|)^{2s} |\mathcal{F}f(u)|^2 du \Big)^{\frac{1}{2}},$$

where $\mathcal{F}f(u) = \int_{\mathbb{R}^d} f(x) e^{i\langle u,x\rangle} dx$ is the Fourier transform of f. The space

$$H^s(\mathbb{R}^d) = \Big\{ f \in L^2(\mathbb{R}^d, \mu) : \|f\|_{H^s} < \infty \Big\}$$

equipped with the norm $\|\cdot\|_{H^s}$ is a real Hilbert space called the Sobolev space of order s.

Inequalities (5.6) and (5.7) assert that

$$\mathbb{E}\left[\left| \Gamma_T(f) - \widehat{\Gamma}_{T,n}(f) \right|^2 \right]^{\frac{1}{2}} \leq \frac{C_T}{n} \left(\|f\|^2_{L^2(\mu)} + \frac{1}{2} \|\nabla f\|^2_{L^2(\mu)} \right)^{1/2} \leq \frac{C_T}{n} \|f\|_{H^1}, \tag{5.8}$$

where we used Plancherel's theorem in the last line. Since obviously

$$\mathbb{E}\left[\left| \Gamma_T(f) - \widehat{\Gamma}_{T,n}(f) \right|^2 \right]^{\frac{1}{2}} \leq \frac{C_T}{\sqrt{n}} \|f\|_{L^2(\mu)} \leq \frac{C_T}{\sqrt{n}} \|f\|_{L^2}, \tag{5.9}$$

by an interpolation argument we can obtain convergence rates that depend on the Sobolev smoothness of the function f.

Corollary 5.7. *Let X be a stationary $d-$dimensional Ornstein-Uhlenbeck process. There exists a constant $C_T < \infty$ such that for any $0 \leq s \leq 1$ and $f \in H^s(\mathbb{R}^d)$ it holds*

$$\mathbb{E}\left[\left| \Gamma_T(f) - \widehat{\Gamma}_{T,n}(f) \right|^2 \right]^{\frac{1}{2}} \leq C_T n^{-\frac{1+s}{2}} \|f\|_{H^s}.$$

Proof. Fix $0 \leq s \leq 1$ and $f \in H^s(\mathbb{R}^d)$. Define the approximating functions

$$f_m = \mathcal{F}^{-1} \left(\mathcal{F}f(\cdot) \mathbf{1}(|\cdot| < m) \right).$$

From inequality (5.8) we deduce

$$\begin{aligned} \mathbb{E}\left[\left| \Gamma_T(f_m) - \widehat{\Gamma}_{T,n}(f_m) \right|^2 \right]^{\frac{1}{2}} &\leq \frac{C_T}{n} \|f\|_{H^1} = \frac{C_T}{(2\pi)^{d/2} n} \Big(\int_{\{|u|<m\}} |\mathcal{F}f_m(u)|^2 (1+|u|)^2 du \Big)^{\frac{1}{2}} \\ &\leq \frac{C_T}{n} (1+m)^{1-s} \|f\|_{H^s}. \end{aligned}$$

Similarly, inequality (5.9) implies

$$\mathbb{E}\left[\left| \Gamma_T(f_m) - \widehat{\Gamma}_{T,n}(f_m) \right|^2 \right]^{\frac{1}{2}} \leq \frac{C_T}{(2\pi)^{d/2} n^{1/2}} \Big(\int_{\{|u|\geq m\}} (\mathcal{F}f_m(u))^2 du \Big)^{\frac{1}{2}} \leq \frac{C_T}{n^{1/2}(1+m)^s} \|f\|_{H^s}.$$

The claim follows by choosing $(1+m) = \sqrt{n}$. ◻

5.4.2. Stationary scalar diffusion processes

Assume that the diffusion coefficients b, σ satisfy the local growth condition:

$$\exists K > 0 \, \forall x \in \mathbb{R} \quad |b(x)| + |\sigma(x)| \le K(1 + |x|),$$

and that the volatility is not degenerated with squared inverse locally integrable:

$$\forall x \in \mathbb{R} \quad \sigma(x) > 0 \quad \text{and} \quad \sigma^{-2} \in L^1_{loc}(\mathbb{R}).$$

Then, for every initial condition X_0 stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{5.10}$$

where W is a Brownian motion, has a unique in the sense of the probability law non-exploding weak solution, see [54, Chapter 5, Theorem 5.15 and Remark 5.19]. Assume in addition that function

$$m(x) = \sigma^{-2}(x) \exp\Big(\int_0^x \frac{2b(y)}{\sigma^2(y)} dy \Big), \quad x \in \mathbb{R}$$

is integrable. Then, equation (5.10) has a stationary solution X (see [46, remark before Assumption 2]) with the invariant density:

$$\frac{d\mu}{dx}(x) = \frac{m(x)}{\int_{-\infty}^{\infty} m(x)dx}. \tag{5.11}$$

The infinitesimal generator L of X is defined by

$$Lf = \mu^{-1}(x) \left(\frac{1}{2} \sigma^2(x) \mu(x) f'(x) \right)', \tag{5.12}$$

with the domain (see Hansen et al. [46, Section 3.3])

$$\begin{aligned} \mathrm{dom}(L) = \Big\{ f \in L^2(\mu) : f' &\text{ exists and is absolutely continuous with} \\ &\lim_{x \to -\infty} f'(x) = \lim_{x \to +\infty} f'(x) = 0 \text{ and } f'' \in L^2(\mu) \Big\}. \end{aligned}$$

For $f, g \in \mathrm{dom}(L)$, integrating by parts, we obtain

$$\langle f, Lg \rangle_\mu = - \int_{\mathbb{R}} f'(x) g'(x) \sigma^2(x) \mu(x) dx.$$

Hence, L is a non-positive self-adjoint operator on $L^2(\mathbb{R}, \mu)$. As such it satisfies the assumptions of Theorem 5.4. Furthermore, arguing like in the case of the Ornstein-Uhlenbeck process the space $\mathcal{D}^1$ can be identified as the space of weakly differentiable functions on $\mathbb{R}$ with the derivative belonging to $L^2(\mathbb{R}^d, \mu)$.

5.4.3. Scalar diffusion process with two reflecting barriers

For a bounded measurable drift $b : [0,1] \to \mathbb{R}$ and continuous volatility $\sigma : [0,1] \to \mathbb{R}_+$ that satisfies

$$\inf_{x\in[0,1]} \sigma(x) > 0$$

consider the following Skorokhod type stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + dK_t,$$

where $(W_t, t \geq 0)$ is a standard Brownian motion and $(K_t, t \geq 0)$ is some adapted continuous process with finite variation, starting form 0, and such that for every $t \geq 0$ it holds

$$\int_0^t \mathbf{1}_{(0,1)}(X_s)dK_s = 0.$$

Furthermore, we assume that the initial distribution X_0 is independent of the driving Brownian motion W and is absolutely continuous with respect to the Lebesgue measure on $[0,1]$ with density

$$\frac{d\mu}{dx}(x) = C_0\sigma^{-2}(x)\exp\Big(\int_0^x 2b(y)\sigma^{-2}(y)dy\Big), \quad x \in [0,1] \tag{5.13}$$

where C_0 is a normalizing constant. The existence and properties of the process X are discussed in Chapter 2. In particular, X is a stationary Markov process induced by a self-adjoint infinitesimal generator

$$Lf(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x), \tag{5.14}$$

with domain

$$\begin{aligned}\mathrm{dom}(L) = \Big\{f \in L^2([0,1]) : f &\text{ has two weak derivatives}\\ &\text{with } f'(0) = f'(1) = 0 \text{ and } f'' \in L^2([0,1])\Big\}.\end{aligned}$$

As such, X satisfies the assumptions of Theorem 5.4. Using integration by parts, together with the so-called divergence form of the generator (see Chapter 2, equation 2.10), one can show that for $f \in \mathrm{dom}(L)$ it holds

$$\|(I-L)^{1/2}f\|^2_{L^2(\mu)} = \|f\|^2_{L^2(\mu)} + \|\frac{1}{2}f'\sigma\|^2_{L^2(\mu)}.$$

Consider $f \in H^1(\mathbb{R})$. Then, by the equivalent definition of the $H^1(\mathbb{R})$ in terms of the weakly differentiable functions with squared integrable derivative it holds that

$$\begin{aligned}\|(I-L)^{1/2}f\|^2_{L^2(\mu)} &\leq \|\mu\|_\infty\|f\|^2_{L^2([L,R])} + \|\sigma^2\|_\infty\|\mu\|_\infty\|f'\|^2_{L^2([L,R])}.\\ &\leq \|\mu\|_\infty(1+\|\sigma^2\|_\infty)\|f\|^2_{H^1}.\end{aligned}$$

Arguing similarly as in the proof of Corollary (5.7), we obtain

Corollary 5.8. *There exists a constant $C < \infty$ such that for any $0 \leq s \leq 1$ and $f \in H^s([0,1])$*

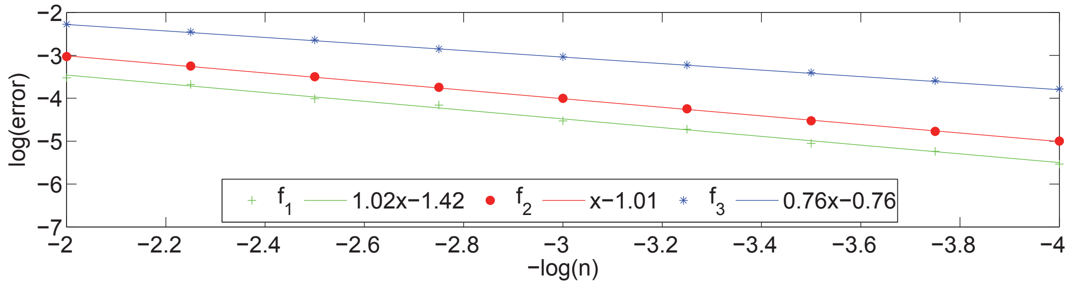

Figure 5.1.: LogLog estimation errors of the occupation times for functions f_1, f_2 and f_3. Solid lines are calculated by linear regression. The empirical convergence rates (slope of the linear regression lines, displayed in the legend) match the theoretical upper bounds.

it holds

$$\mathbb{E}\left[\left|\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)\right|^2\right]^{\frac{1}{2}} \leq (1 + \|\mu\|_\infty)(1 + \|\sigma^2\|_\infty)\sqrt{1 \vee T} n^{-\frac{1+s}{2}} \|f\|_{H^s}. \tag{5.15}$$

5.4.4. Numerical study

In this section we present numerical results for the estimation of the occupation time. Let X be a stationary, one dimensional Ornstein-Uhlenbeck process, as defined in Section 5.4.1. Note that since X is a Gaussian process with known covariance structure, we can sample the observations $X_0, X_{\frac{1}{n}}, ..., X_{\frac{\lfloor nT \rfloor}{n}}$ without error.

In Figure 5.1, we present the loglog plot of the mean estimation errors

$$\mathbb{E}\left[\left|\Gamma_T(f) - \widehat{\Gamma}_{T,n}(f)\right|^2\right]^{\frac{1}{2}}$$

approximated by independent Monte Carlo simulation with 1000 iterations. We present the results for three functions with different Sobolev regularity:

- the sinc function $f_1(x) = \frac{\sin(x)}{x}$, $f_1 \in H^s(\mathbb{R})$, for all $0 < s < \infty$
- function f_2 satisfying $\mathcal{F} f_2(u) = \frac{1}{(1+|u|)^{3/2}}$, such that $f_2 \in H^s$ for $0 < s < 1$
- indicator function $f_3(x) = \mathbf{1}_{[0,1]}(x)$, $f_3 \in H^s$ for $0 < s < 1/2$.

In order to compute values of f_2 we inverted numerically the Fourier transform as described in Tankov [88, Section 11.1.3 p. 366]. Using Corollary 5.8 and the Sobolev smoothness of the considered functions we obtain convergence rates n^{-1} for f_1, f_2 and $n^{-(3/4-\varepsilon)}$, $\varepsilon > 0$ for f_3. As demonstrated in Figure 5.1, the empirical convergence rates, calculated as the slope of the linear fit, match the theoretical rates very well.

6. Conclusion

The inference problem for the coefficients of a time-homogeneous Itô diffusion has been studied intensively and is now well-understood. Nevertheless, existing methods are designed under specific assumptions about the sampling schemes of observations. The locally averaged spectral estimator $\widehat{\sigma}_S$, presented in Chapter 4, is probably the first known method that adapts effectively to the sampling frequency. Its greatest advantage is that it leaves out the question of the right frequency paradigm when one has to consider data. Its biggest flaw is that its performance does not improve when σ is more than once differentiable. For a practitioner it could at least provide intuition about the shape of the volatility function before referring to frequency specific methods.

After establishing that $\widehat{\sigma}_S$ reaches optimal rates in both low and high-frequency regimes, it is natural to ask about the joint dependence of the error on both sampling parameters Δ and T. This problem is equally interesting as difficult and requires a new strategy of the proof. The proofs of the convergence rates presented in Chapter 4 require in high and low-frequency regimes that T and Δ are fixed. This is especially noticeable in the proof of the low frequency convergence rate, where it is crucial that the spectral gap (and hence Δ) has a positive lower bound.

To overcome technical complexity due to structural differences between low and high-frequency data the model assumptions are quite specific. Since the stationarity condition is used mainly to ensure the time-reversibility of the diffusion, it could probably be replaced by the assumption that the coefficients of the time reversed process are regular enough. For applications in empirical econometrics or finance, the most restrictive assumption is the boundary reflection. When the time horizon of the observations is fixed, the volatility inference is possible in a compact domain only. Hence, the estimation error of the eigenpair $(\widehat{\xi}_1, \widehat{u}_1)$ would have to be controlled locally, which, if the state space is unbounded, is very difficult due to the global character of the a posteriori error bounds on solutions of the eigenvalue problems.

In a recent paper, Chen et al. [20] discussed the existence and extraction of eigenpairs of generators of multivariate diffusions. Given the existence of a spectral gap of the generator, same methods as in Chapter 4 can be applied to estimate the eigenpairs of the generator in a multidimensional setting. Nevertheless, to propose a consistent statistical program one needs to find the multivariate equivalent of the spectral representation of the diffusion coefficients in terms of the invariant density and the eigenpairs of the generator. The generalization of the one-dimensional formula is not straightforward, since it is based on the integral representation of the antiderivative, which does not have a multivariate counterpart. Under certain symmetry conditions the volatility could be inferred from a system of equations involving the first d eigenpairs (where d is the dimension of the considered diffusion), which imposes further conditions on the eigenvalues of the generator.

Chapter 5 is devoted to the problem of inference of the occupation time functional for non-continuous functions. This intriguing problem was considered only recently and still waits for a comprehensive analysis. Results presented in this thesis are part of an ongoing research project with Randolf Altmeyer. In the setting of Markov processes an interesting question is the exact role of the stationarity assumption, especially that it has to be considered before relaxing the stationarity assumption on the frequency universal estimator. When X is an Itô semimartingale, using Fourier analysis one can obtain central limit theorems and consequently confidence bounds for the Riemann estimator.

A. Stability of the eigenvalue problems

A.1. Compact, positive-definite, self adjoint operator

Theorem A.1. Consider T a compact, self-adjoint and positive-definite operator on some Hilbert space $\mathcal{H} = (H, \|\cdot\|)$. Denote its eigenpairs by $(\lambda_i, x_i)_{i=1,2,\dots}$, normalized so that $\|x_i\| = 1$ and ordered decreasingly with respect to the eigenvalues. Let $V \subset H$ be a finite dimensional subspace of H, and π the orthogonal projection on V. Assume that the biggest eigenvalue λ_1 is simple and that

$$\|(I-\pi)\, x_1\| < \frac{\lambda_1 - \lambda_2}{6\lambda_1}.$$

Consider the projected operator $\pi T \pi$ and denote its normalized, ordered decreasingly, eigenpairs by $(\lambda_{V,i}, x_{V,i})_{i=1,2,\dots,\dim(V_J)}$. Then $\lambda_{V,1} - \lambda_{V,2} \geq (\lambda_1 - \lambda_2)/2$ and

$$|\lambda_1 - \lambda_{V,1}| + \|x_1 - x_{V,1}\| \leq C \, \|(I-\pi)\, x_1\| \,,$$

where the constant C depends continuously only on the size of the spectral gap $\lambda_1 - \lambda_2$ and the value of the first eigenvalue λ_1.

Proof. Since T is self-adjoint and positive-definite

$$\|T\| = \sup_{x \in H} \frac{\langle Tx, x\rangle}{\|x\|^2} = \lambda_1.$$

By the variational characterization of the eigenvalues

$$\lambda_{V,i} = \sup_{\substack{S \subset V \\ \dim(S)=i}} \inf_{y \in S} \frac{\langle y, Ty\rangle}{\|y\|^2} \leq \sup_{\substack{S \subset H \\ \dim(S)=i}} \inf_{y \in S} \frac{\langle y, Ty\rangle}{\|y\|^2} = \lambda_i. \tag{A.1}$$

Furthermore,

$$\begin{aligned}
\lambda_1 - \lambda_{V,1} &\leq \frac{\langle (\lambda_1 - \pi T \pi)\,(\pi x_1)\,, \pi x_1\rangle}{\|\pi x_1\|^2} = \frac{\langle \pi T\,(I-\pi)\, x_1, \pi x_1\rangle}{\|\pi x_1\|^2} \\
&\leq \frac{\|\pi T\,(I-\pi)\, x_1\|}{\|\pi x_1\|} \leq \|T\| \frac{\|(I-\pi)\, x_1\|}{\|\pi x_1\|} \\
&\leq \|T\| \frac{\|(I-\pi)\, x_1\|}{1 - \|(I-\pi)\, x_1\|}.
\end{aligned}$$

Since $|\lambda_1 - \lambda_{V,1}| \leq 2\|T\|$, from the inequality $\frac{z}{1-z} \wedge 2 \leq 3z$ for $z = \|(I-\pi)\, x_1\|$ follows that

$$|\lambda_1 - \lambda_{V,1}| \leq 3\|T\| \, \|(I-\pi)\, x_1\| \,.$$

Since by (A.1) holds $\lambda_{V,2} \leq \lambda_2$ and $\|T\| \, \|(I-\pi)\, x_1\| < \frac{\lambda_1 - \lambda_2}{6}$ we have

$$\begin{aligned} |\lambda_{V,1} - \lambda_{V,2}| &\geq \lambda_{V,1} - \lambda_2 = |\lambda_1 - \lambda_2| - |\lambda_1 - \lambda_{V,1}| \\ &\geq \lambda_1 - \lambda_2 - 3\|T\| \, \|(I-\pi)\, x_1\| \geq \frac{1}{2}\left(\lambda_1 - \lambda_2\right). \end{aligned}$$

Consequently, the projected operator $\pi T \pi$ has a spectral gap of size $\rho \geq \frac{\lambda_1 - \lambda_2}{2}$ and in particular the eigenvalue $\lambda_{V,1}$ is simple. Define the residual vector $r = (\pi T \pi - T)\, x_1$. Then

$$\begin{aligned} \|r\| = \|(\pi T \pi - T)\, x_1\| &\leq \|\pi T \pi x_1 - \pi T x_1\| + \lambda_1 \|\pi x_1 - x_1\| \\ &\leq (\|T\| + \lambda_1)\, \|(I-\pi)\, x_1\|. \end{aligned}$$

We conclude that in order to prove $\|x_1 - x_{V,1}\| \leq C \, \|(I-\pi)\, x_1\|$, it suffices to justify that

$$\|x_1 - x_{V,1}\| \leq \frac{3\rho^2}{2\sqrt{2}} \, \|r\|$$

Let P be the spectral projection on the eigenspace of operator $\pi T \pi$ corresponding to the eigenvalue λ_1^V. Let $R(\pi T \pi, z) = (\pi T \pi - z)^{-1}$ be the resolvent operator. Using Cauchy's integral representation of the spectral projection (see Lemma 6.4 from [19]) together with $|\lambda_1 - \lambda_{V,1}| \leq \rho$ we find

$$\begin{aligned} \|x_1 - P x_1\| &= \frac{1}{2\pi} \Big\| \oint_{S(\lambda_1, 3\rho/2)} \frac{R(\pi T \pi, z)}{\lambda_1 - z} dz \, (\pi T \pi - T)\, x_1 \Big\| \\ &\leq \frac{3\rho}{2} \|r\| \sup_{z \in S(\lambda_1, 3\rho/2)} \|R(\pi T \pi, z)\|. \end{aligned}$$

Since operator $\pi T \pi$ is self adjoint on $\mathcal{H}$, we know that (see Proposition 2.32 from [19]) $\|R(\pi T \pi, z)\| = (\operatorname{dist}(z, \sigma(\pi T \pi)))^{-1}$. Hence

$$\sup_{z \in S(\lambda_1, 3\rho/2)} \|R(\pi T \pi, z)\| = \sup_{z \in S(\lambda_1, 3\rho/2)} (\operatorname{dist}(z, \sigma(\pi T \pi)))^{-1} \leq \frac{\rho}{2}.$$

It remains to bound the distance between the eigenvectors. Vectors x_1 and $x_{V,1}$ are normalized, hence

$$\begin{aligned} \|x_{V,1} - x_1\|^2 &= 2 - 2\langle x_{V,1}, x_1\rangle \leq 2 - 2\langle x_{V,1}, x_1\rangle^2 \\ &= 2\left(1 + \langle x_{V,1}, x_1\rangle\right)\left(1 - \langle xV_{,1}, x_1\rangle\right) = 2 \, \|x_1 - \langle x_{V,1}, x_1\rangle x_{V,1}\|^2. \end{aligned}$$

As $\lambda_{V,1}$ is simple, the right hand side is equal to $2 \, \|x_1 - P x_1\|^2$. □

A.2. Bilinear coercive form

For differentiable, strictly positive functions σ and μ consider an elliptic operator T on $L^2([0,1])$, with Neumann type domain $\operatorname{dom}(T) = \{v \in H^2 : v'(0) = v'(1) = 0\}$, given in the divergence form by

$$Tv(x) = -\frac{(\sigma^2(x)\mu(x)v'(x))'}{2\mu(x)}, \text{ for } v \in \mathrm{dom}(T). \tag{A.2}$$

Note that the operator $-T$ is an infinitesimal generator of the diffusion process on $[0,1]$ with instantaneous reflection at the boundaries, volatility function σ and an invariant measure with density μ with respect to the Lebesgue measure. We want to analyze the eigenvalue problem for T, i.e.

Eigenproblem A.2. Find $(\lambda, w) \in \mathbb{R} \times \mathrm{dom}(T)$, with $w \neq 0$, such that

$$Tw = \lambda w.$$

Integrating by parts, one can check, that the eigenpairs of the Eigenproblem A.2 solve

Eigenproblem A.3. Find $(\lambda, w) \in \mathbb{R} \times H^1$, with $w \neq 0$, such that

$$\int_0^1 w'(x)v'(x)\sigma^2(x)\mu(x)dx = 2\lambda \int_0^1 w(x)v(x)\mu(x)dx \text{ for all } v \in H^1. \tag{A.3}$$

Eigenproblem A.3 is a weak formulation of the Eigenproblem A.2 for the associated Dirichlet form $l(u,v) = \langle Tu, v\rangle_\mu$. The biggest advantage of the weak formulation is that the Eigenproblem A.3 makes sense for any, not necessarily regular, functions μ. When μ is not differentiable, the Eigenproblem A.2 has no longer probabilistic interpretation in terms of the infinitesimal generator. Nevertheless, such problems arise naturally when one considers spectral estimation method with fixed time horizon, when the role of the invariant measure is taken by the non differentiable occupation density.

In what follows, we want to generalize the results of [44] on the spectral properties of an infinitesimal generator, to the solutions of the Eigenproblem A.3 with a Hölder regular function μ.

Definition A.4. For any given $0 < d < D$ let

$$\Theta_\alpha := \Big\{(\sigma,\mu) \in H^1([0,1]) \times C^{0,\alpha}([0,1]) : \|\sigma\|_{H^1}, \|\mu\|_{C^{0,\alpha}} \leq D,$$
$$\inf_{x\in[0,1]}(\sigma(x) \wedge \mu(x)) \geq d, \int_0^1 \mu(x)dx = 1\Big\}$$

Eigenproblem A.3 is a conforming eigenvalue problem for a bilinear coercive form on the Hilbert space $L^2(\mu)$. [19] is a standard reference.

Proposition A.5. *Let $(\sigma,\mu) \in \Theta_\alpha$. The Eigenproblem A.3 has countably many solutions $(\lambda_i, w_i)_i$, with real nonnegative eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots$ and $\mu-$orthogonal eigenfunctions, satisfying Neumann boundary conditions $w_i'(0) = w_i'(1) = 0$. The smallest positive eigenvalue λ_1 is simple, the corresponding eigenfunction $w_1 \in C^{1,1/2\wedge\alpha}$ and is strictly monotone.*

Proof. It is easy to check that for any (σ,μ) $\lambda_0 = 0$ and $w_0 \equiv 1$ form an eigenpair. Let $L_0^2(\mu) = \{v \in L^2(\mu) : \int_0^1 v(x)\mu(x)dx = 0\}$ and $H_0^1(\mu) = L_0^2(\mu) \cap H^1$. $L_0^2(\mu)$ with the $L^2(\mu)$

inner product and $H_0^1(\mu)$ with $\langle u,v\rangle_{H^1(\mu)} = \langle u,v\rangle_{L_0^2(\mu)} + \int_0^1 u'(x)v'(x)\mu(x)dx$ are Hilbert spaces. The identity embedding $I : H_0^1(\mu) \to L_0^2(\mu)$ is compact.

For $u, v \in H_0^1(\mu)$ let

$$l(u,v) = \int_0^1 u'(x)v'(x)\sigma^2(x)\mu(x)dx.$$

l is a symmetric positive-definite bilinear form on $H_0^1(\mu) \times H_0^1(\mu)$. Furthermore, for any $u \in H_0^1(\mu)$ holds

$$c\|u\|^2_{H_0^1(\mu)} \le l(u,u) \le C\|u\|^2_{H_0^1(\mu)}, \tag{A.4}$$

for some constants $0 < c < C$ that depend only on d, D. Indeed, since σ and μ are uniformly bounded, we only have to show that $\int_0^1 u^2(x)dx \le \int_0^1 (u'(x))^2 dx$. Consider $u \in C^1([0,1]) \cap H_0^1(\mu)$. Since u is continuous and integrates to zero, there exists $x_0 \in [0,1]$ s.t. $u(x_0) = 0$. Since $u(x) = \int_{x_0}^x u'(y)dy$, the upper bound $\|u\|_{L^2} \le \|u'\|_{L^2}$ follows from the Cauchy-Schwarz inequality. As Lipschitz functions are dense in H^1, we conclude that (A.4) holds.

l is the Dirichlet form of an unbounded operator T on $L_0^2(\mu)$. Define $D = \mathrm{dom}(T)$ as these $u \in H_0^1(\mu)$, that the functional $v \mapsto l(u,v)$ is continuous on $H_0^1(\mu)$ with norm $\|\cdot\|_{L^2(\mu)}$. By the definition of the weak differentiability, domain $D = \{u : H_0^1(\mu) : u'\sigma^2\mu \in H^1\}$. Furthermore, D is dense in $L_0^2(\mu)$ (see [19, Exercise 4.51]). For $u \in D$, we define Tu via the Riesz representation theorem by $l(u,v) = \langle Tu, v\rangle_{L^2(\mu)}$. Such defined T is an elliptic, densely defined, self-adjoint operator with compact resolvent (see [19, Proposition 4.17]). Consequently, T has a discrete spectrum $(\lambda_i)_{i=1,\ldots}$, with all eigenvalues positive and corresponding eigenfunctions $\mu-$orthogonal.

Integrating by parts the right hand side of (A.3), we obtain

$$\int_0^1 w_i'(x)\sigma^2(x)\mu(x)v'(x)dx = -2\lambda_i \int_0^1 \int_0^x w_i(y)\mu(y)dyv'(x)dx \text{ for all } v \in H^1.$$

Since $\{v' : v \in H^1\}$ is dense in L^2, it follows that

$$w_i'(x) = \frac{2\lambda_i \int_0^x w_i(y)\mu(y)dy}{\sigma^2(x)\mu(x)}. \tag{A.5}$$

By Sobolev embedding σ^2 is $1/2-$Hölder regular. Consequently $w_i' \in C^{0,1/2\wedge\alpha}$. Since the eigenfunctions $\mu-$integrate to zero, we deduce that $w_i'(0) = w_i'(1) = 0$.

Finally, we need to show that λ_1 is simple and that w_1 is strictly monotone. By the variational formula for the eigenpairs of a self-adjoint operator

$$2\lambda_1 = \inf_{u \in H_0^1(\mu)} \frac{\int_0^1 (u'(x))^2\sigma^2(x)\mu(x)dx}{\int_0^1 u^2(x)\mu(x)dx}. \tag{A.6}$$

Arguing as in [44, Lemma 6.1], we obtain that

$$\int_0^1 u^2(x)\mu(x)dx = \int_0^1\int_0^1 m(y,z)u'(y)u'(z)dydz,$$

with

$$m(y,z) = \int_0^{y\wedge z} \mu(x)dx \int_{y\vee z}^1 \mu(x)dx.$$

We deduce that the derivative of the eigenfunction w_1 must have a constant sign, otherwise we could reduce the ratio in (A.6) by considering

$$\widetilde{w}_1 = w_1 \mathbf{1}(w_1' \geq 0) - w_1 \mathbf{1}(w_1' \leq 0).$$

Hence, the set $\{x : w_1'(x) = 0\}$ has zero Lebesgue measure. From (A.5) follows that $w_1'(x) = 0$ only for $x = 0, 1$, meaning that w_1 is strictly monotone on $(0,1)$. Consequently, for any two eigenfunctions w_1 and $\bar{w}_1$, which correspond to λ_1, the scalar product

$$\int_0^1 w_1(x)\bar{w}_1(x)\mu(x)dx = \int_0^1 \int_0^1 m(y,z) w_1'(y)\bar{w}_1'(z)dydz \neq 0,$$

hence the eigenspace corresponding to λ_1 is one dimensional. □

Proposition A.6. *The eigenvalues λ_1, λ_2 and the norm ratio $\|w_1\|_{C^{1,1/2\wedge\alpha}}/\|w_1\|_{L^2(\mu)}$ are uniformly bounded for all $(\sigma,\mu) \in \Theta_\alpha$. Furthermore, for every $0 < a < b < 1$, $\inf_{x\in[a,b]} |w_1'(x)|$ and the spectral gap $\lambda_2 - \lambda_1$ have uniform lower bounds on Θ_α.*

Proof. We adapt the notation from the proof of Proposition A.5. Choose w_1 normalized s.t. $\|w_1\|_{L^2(\mu)} = 1$. We will first argue that λ_1, λ_2 and $\|w_1\|_{C^{1,1/2\wedge\alpha}}$ are uniformly bounded on Θ_α. From (A.4) we imply that w

$$2\lambda_1 = l(w_1, w_1) \geq c\|w_1\|^2_{H^1(\mu)} \geq c,$$

with $c > 0$ depending only on the bounds on σ and μ. It follows that the eigenvalues are uniformly separated from zero. By the variational formula

$$2\lambda_2 = \inf_{\substack{S\subset H^1\\ \dim(S)=3}} \sup_{u\in S} \frac{\int_0^1 (u'(x))^2\sigma^2(x)\mu(x)dx}{\int_0^1 u^2(x)\mu(x)dx} \leq \inf_{\substack{S\subset H^1\\ \dim(S)=3}} \sup_{u\in S} \frac{D^3\int_0^1 (u'(x))^2 dx}{d\int_0^1 u^2(x)dx} \leq 4\pi^2\frac{D^3}{d},$$

since $4\pi^2$ is the third eigenvalue of the negative Laplace operator on $L^2([0,1])$ with Neumann boundary conditions. We conclude that the eigenvalues λ_1 and λ_2 are uniformly bounded. The uniform bound on $\|w_1\|_{C^{1,1/2\wedge\alpha}}$ follows from the representation (A.5).

We will now prove a uniform lower bound on the spectral gap $\lambda_2 - \lambda_1$. Assume by contradiction that for some sequence of coefficients $(\sigma_n, \mu_n) \in \Theta_\alpha$ the corresponding spectral gaps $(\lambda_{n,2} - \lambda_{n,1})$ converge to zero. Since Θ_α is compact in the uniform convergence metric, we can assume that (σ_n, μ_n) converges uniformly to some $(\sigma,\mu) \in \Theta_\alpha$. We will argue that the uniform convergence of the coefficients leads to convergence of the eigenvalues, hence contradicts Proposition A.5 (cf. [44, proof of Proposition 6.5]). However, since the function μ is embedded in the definition of spaces $L_0^2(\mu)$ and $H_0^1(\mu)$, we need first to reduce the Eigenproblem A.3 to a universal function space.

Let $U(x) = \int_0^x \mu(y)dy$ be the distribution function of μ. Substituting $U(x) = y$, we find that the Eigenproblem A.3 is equivalent to

$$\begin{aligned} \int_0^1 \widetilde{w}'(x)\widetilde{v}'(x)\widetilde{\sigma}^2 dx &= 2\lambda \int_0^1 \widetilde{w}'(x)\widetilde{v}'(x)dx \text{ for all } \widetilde{v} \in H^1 \\ \widetilde{w} &= w \circ U^{-1}, \end{aligned}$$

with $\widetilde{\sigma} = (\sigma\mu) \circ U^{-1}$. Consider $(\widetilde{\sigma}_n)_n$ and $\widetilde{\sigma}$ corresponding to (σ_n, μ_n) and (σ, μ) respectively. Note that $\widetilde{\sigma}_n$ converges to $\widetilde{\sigma}$ in the uniform norm. Denote $L_0^2 = L_0^2(1)$ and $H_0^1 := H_0^1(1)$. For $u, v \in H_0^1$ denote

$$\widetilde{l}_n(u,v) = \int_0^1 u'(x)v'(x)\widetilde{\sigma}_n(x)^2 dx$$

and by $\widetilde{T}_n$ the corresponding operators on L_0^2. Recall that the operators $\widetilde{T}_n$ are unbounded and self-adjoint on L_0^2, with dense domains $\widetilde{D}_n$. Domains $\widetilde{D}_n$ do not have to possess a common core, which is needed to study the convergence of the sequence $(\widetilde{T}_n)_n$. We circumvent this difficulty by introducing inverse operators $\widetilde{R}_n = \widetilde{T}_n^{-1}$. Using the divergence formula (A.2) for $\widetilde{T}_n$, we check that for $u \in L_0^2$

$$\widetilde{R}_n u(x) = -2\int_0^x \widetilde{\sigma}_n^{-2}(y) \int_0^y u(z)dz + c_n(u), \tag{A.7}$$

where $c_n(u) \in \mathbb{R}$ is such that $\int_0^1 \widetilde{R}_n u(x)dx = 0$. The convergence $\widetilde{\sigma}_n \to \widetilde{\sigma}$ in $C^1[(0,1)]$ implies that operators $\widetilde{R}_n$ converge to $\widetilde{R}$ in the operator norm on L_0^2. By [19, Proposition 5.28] this entails the regular convergence, which, by [19, Theorem 5.20], is equivalent to the strongly stable convergence. Finally, [19, Proposition 5.6] ensures the convergence of the eigenvalues with preservation of their multiplicities.

Set $0 < a < b < 1$. We finally have to prove the uniform lower bound on $\inf_{x\in[a,b]} |w_1'(x)|$. We will use the same indirect arguments as when bounding the spectral gap. Assume that for some sequence $(\sigma_n, \mu_n) \in \Theta_\alpha$, with (σ_n, μ_n) converging in the uniform norm to $(\sigma, \mu) \in \Theta_\alpha$, the corresponding eigenfunctions $w_{1,n}$ satisfy $\inf_n \inf_{x\in[a,b]} |w_{1,n}'(x)| = 0$. Arguing as for the spectral gap, we reduce the problem to bounded operators $(\widetilde{R}_n)_n$ and $\widetilde{R}$. From formula (A.7) we deduce that the uniform convergence of coefficients implies $\widetilde{R}_n \to \widetilde{R}$ in the operator norm on $C([0,1])$. We conclude, that the eigenfunctions converge in the uniform norm, which contradicts Proposition A.5. □

Eigenproblem A.7. Let V_J be a finite dimensional subspace of L^2. Find $(\lambda_J, w_J) \in \mathbb{R} \times V_J$, with $w_J \neq 0$ such that

$$\int_0^1 w'(x)v'(x)\sigma^2(x)\mu(x)dx = \lambda \int_0^1 w(x)v(x)\mu(x)dx \text{ for any } v \in V_J.$$

Proposition A.8. *Let $(V_J)_{J=1,\dots}$ be a sequence of approximation spaces satisfying the following Jackson's type inequality:*

$$\|(I - \pi_J)v\|_{H^1} \leq CJ^{-\alpha}\|v\|_{C^{1,\alpha}} \textit{ for } v \in C^{1,\alpha},$$

where π_J is the L^2-orthogonal projection on V_J and $C > 0$ some universal constant . Furthermore, assume that every V_J contains constant functions.

For $(\sigma, \mu) \in \Theta_\alpha$ the Eigenproblem A.7 has $\dim(V_J)$ solutions $(\lambda_{J,i}, w_{J,i})_i$ with real eigenvalues $0 = \lambda_{J,0} < \lambda_{J,1} < \lambda_{J,2} \leq \dots \leq \lambda_{J,\dim(V_J)-1}$. For J big enough, the eigenvalue $\lambda_{J,1}$ and the spectral gap $\lambda_{J,2} - \lambda_{J,1}$ are uniformly bounded on Θ_α.

Proof. We adapt the notation from the proof of Proposition A.5. By the Lax-Milgram theorem, there exists an isomorphism $S_l : H_0^1(\mu) \to H_0^1(\mu)$ such that

$$l(S_l v, u) = \langle v, u\rangle_{H^1(\mu)}, \text{ for all } v, u \in H_0^1(\mu).$$

Note that since for any $v \in L_0^2(\mu)$ the functional $H_0^1(\mu) \ni u \longmapsto \langle v, u\rangle_{L^2(\mu)} \in \mathbb{R}$ is continuous on $H_0^1(\mu)$, by the Riesz representation theorem there exists a continuous operator $K : L_0^2(\mu) \to H_0^1(\mu)$ such that

$$\langle v, u\rangle_{L^2(\mu)} = \langle Kv, u\rangle_{H^1(\mu)}.$$

Define the operator $B_l = S_l \circ K \circ I$, where I is the identity embedding of $H_0^1(\mu)$ into $L_0^2(\mu)$. By (A.4), the form l defines an equivalent norm on $H_0^1(\mu)$. Note that B_l is a self-adjoint and compact operator on the Hilbert space $H_0^1(\mu)$ with $l-$induced inner product. Consider (λ_i, w_i), a solution of the Eigenproblem A.3. For any $v \in H_0^1(\mu)$ we have

$$l(w_i, v) = \lambda_i\langle w_i, v\rangle_{L^2(\mu)} = \lambda_i\langle Kw_i, v\rangle_{H^1(\mu)} = \lambda_i l(S_l K w_i, v) = l(\lambda_i B_l w_i, v),$$

hence (λ_i^{-1}, w_i) is an eigenpair of the operator B_l. In particular, Proposition A.5 implies that the biggest eigenvalue λ_1^{-1} is simple.

Denote by π_J^l the $l-$orthogonal projection on the subspace V_J. Define the operator $B_{l,J} = \pi_J^l B_l \pi_J^l$. Since $B_{l,J}$ is a self-adjoint operator on V_J, with the $l-$induced inner product, it has $\dim(V_J) - 1$ solutions $(\lambda_{J,i}^{-1}, w_{J,i})_i$, with the eigenvalues $\lambda_{J,1}^{-1} \geq \lambda_{J,2}^{-1} \geq \ldots \geq \lambda_{J,\dim(V_J)-1}^{-1}$. Analogously as for the operator B_l, we check that $(\lambda_{J,i}, w_{J,i})$ are solutions of the finite dimensional Eigenproblem A.7. From (A.4) together with the uniform bound on μ follows that

$$\|(I - \pi_n^l)w_1\|_l \leq \|(I - \pi_n^l)(I - \pi_J)w_1\|_l \leq 2\|(I - \pi_J)w_1\|_l \leq C\|(I - \pi_J)w_1\|_{H^1},$$

for some, uniform on Θ_α, constant C. Using Jackson's inequality, the uniform bound on the Hölder norm of w_1 and uniform bounds on the eigenvalues λ_1, λ_2, we conclude that, for J large enough,

$$\|(I - \pi_n^l)w_1\|_l < \frac{\lambda_1^{-1} - \lambda_2^{-1}}{6\lambda_1^{-1}}.$$

The claim follows from Theorem A.1. □

A.3. Generalized eigenvalue problem for positive definite symmetric matrix pairs

In this section we want to sketch the a posteriori technique of solving generalized symmetric eigenvalue problems (GSEP). GSEPs have been studied extensively in chapter VI of [84]. For the error analysis in the case of standard matrix eigenvalue problems we refer to Chapter 1 of [19] or Chapter V of [84]. A particularly useful reference for various eigenvalue problems is [7].

Consider $A, B \in \mathbb{R}^{n\times n}$ real, symmetric matrices with B positive definite. We call a pair $(\lambda, x) \in \mathbb{R} \times (\mathbb{R}^n \setminus \{0\})$ an eigenpair of the generalized symmetric eigenvalue problem (GSEP) for matrices A, B if

$$Ax = \lambda Bx. \tag{A.8}$$

Furthermore, we adapt the notation of the standard eigenvalue problems calling λ the eigenvalue and x the eigenvector. An eigenpair is normalized if $\|x\| = 1$, where $\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$ is the Euclidean norm on $\mathbb{R}^n$.

Using the Cholesky decomposition of matrix $B = DD^*$ one can reduce the generalized eigenproblem (A.8) to the standard eigenvalue problem for matrix $D^{-1}AD^{-*}$. We deduce that (A.8) has n solutions $(\lambda_i, x_i)_{i=1,..,n}$, with all eigenvalues real. Thus we can order the eigenpairs with respect to the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. Finally, the corresponding eigenvectors $(x_i)_{i=1,..,n}$ form a $B-$orthogonal basis of $\mathbb{R}^n$.

Consider now the perturbed matrices $\widetilde{A}$, $\widetilde{B}$ with $\widetilde{B}$ positive definite and the corresponding GSEP:

Eigenproblem A.9. Find $(\widetilde{\lambda}, \widetilde{x}) \in \mathbb{R} \times (\mathbb{R}^n \setminus 0)$ such that

$$\widetilde{A}\widetilde{x} = \widetilde{\lambda}\widetilde{B}\widetilde{x}. \tag{A.9}$$

We want to bound the error between the eigenpairs $(\widetilde{\lambda}_1, \widetilde{x}_1)$ and (λ_1, x_1). To that purpose we form the residual vector

$$r = A\widetilde{x}_1 - \widetilde{\lambda}_1 B\widetilde{x}_1 = (A - \widetilde{A})\widetilde{x}_1 + \widetilde{\lambda}_1(\widetilde{B} - B)\widetilde{x}_1.$$

The standard a posteriori procedure is to find a matrix $E = E(\widetilde{\lambda}_1, \widetilde{x}_1)$ such that

$$\begin{aligned} (A + E)\widetilde{x}_1 &= \widetilde{\lambda}_1 B\widetilde{x}_1, \\ \|E\| &= \|r\|. \end{aligned} \tag{A.10}$$

Since we replaced in (A.10) the perturbed matrix $\widetilde{B}$ by B, the final step is to reduce (A.10) and (A.8) to the standard eigenvalue problems using the Cholesky decomposition of B. Then, we can apply classical error bounds expressed in terms of the perturbation matrix E , e.g. [84, Chapter V, Exercise 1] We obtain

Theorem A.10. There exists a normalized eigenpair (λ_i, x_i), $1 \leq i \leq n$ such that

$$\begin{aligned} |\lambda_i - \widetilde{\lambda}_1| &\leq \left\|B^{-1}\right\| \|r\|, \\ \|x_i - \widetilde{x}_1\| &\leq \frac{2\sqrt{2\kappa(B)}}{\delta(\lambda_i)} \left\|B^{-1}\right\| \|r\|. \end{aligned}$$

where $\kappa(B) = \|B\|\|B^{-1}\|$is the condition number of matrix B and $\delta(\lambda_i)$ is the so called localizing distance, i.e. $\delta(\lambda_i) = \min_{j \neq i} \left|\lambda_j - \widetilde{\lambda}_1\right|$.

The disadvantage of the above procedure is that we obtain an existence result that gives no information how the eigenpair (λ_i, x_i) is related to (λ_1, x_1). This is a typical downside for a posteriori methods that are supposed to provide information how far the calculated solution is from the nearest exact solution but are not intended to compare ordered eigenpairs. A helpful result is the absolute Weyl theorem for generalized hermitian definite matrix pairs, established by Y. Nakatsukasa [62]. For readers convenience we state below the theorem in the form presented in [63, Theorem 8.3].

Theorem A.11. Let $\lambda_1 \geq ... \geq \lambda_n$ and $\widetilde{\lambda}_1 \geq ... \geq \widetilde{\lambda}_n$ be respectively exact and approximated eigenvalues of problems (A.8) and (A.9). Denote $\Delta A = A - \widetilde{A}$ and $\Delta B = B - \widetilde{B}$. Then

$$\begin{aligned}\left|\lambda_i - \widetilde{\lambda}_i\right| &\leq \left\|\widetilde{B}^{-1}\right\| \left\|\Delta A - \lambda_i \Delta B\right\|, \\ \left|\lambda_i - \widetilde{\lambda}_i\right| &\leq \left\|B^{-1}\right\| \left\|\Delta A - \widetilde{\lambda}_i \Delta B\right\|,\end{aligned}$$

for all $i = 1, ..., n$.

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 9th edition.

[2] Adamczak, R. (2007). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab*, 13:1000–1034.

[3] Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, 70(1):223–262.

[4] Aït-Sahalia, Y. (2010). *Econometrics of Diffusion Models.* John Wiley and Sons, Ltd.

[5] Aït-Sahalia, Y. and Mykland, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, 71(2):483–549.

[6] Aït-Sahalia, Y. and Mykland, P. A. (2004). Estimators of diffusions with randomly spaced discrete observations: a general theory. *The Annals of Statistics*, 32(5):2186–2222.

[7] Bai, Z. and Li, R. (2000). Stability and accuracy assessments. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide.*

[8] Ball, C. A. and Roma, A. (1998). Detecting mean reversion within reflecting barriers: application to the European Exchange Rate Mechanism. *Applied Mathematical Finance*, 5:1–15.

[9] Bandi, F. and Phillips, P. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283.

[10] Banon, G. (1978). Nonparametric Identification for Diffusion Processes. *SIAM Journal on Control and Optimization*, 16(3):380–395.

[11] Baraud, Y. (2010). A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli*, 16(4):1064–1085.

[12] Bass, R. (1995). *Probabilistic Techniques in Analysis.* Springer-Verlag, New York.

[13] Bass, R. (1997). *Diffusions and Elliptic Operators.* Springer-Verlag, New York.

[14] Bibby, B. M. and Sørensen, M. (1995). Martingale Estimation Functions for Discretely Observed Diffusion Processes. *Bernoulli*, 1(1/2):17–39.

[15] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes - Estimation and Prediction.* Springer, New York, 2 edition.

[16] Brown, B. M. and Hewitt, J. I. (1975). Asymptotic Likelihood Theory for Diffusion Processes. *Journal of Applied Probability*, 12(2):228–238.

[17] Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine Series 2*, 4(21):161–173.

[18] Chapman, S. (1928). On the Brownian Displacements and Thermal Diffusion of Grains Suspended in a Non-Uniform Fluid. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 119(781):34–54.

[19] Chatelin, F. (1983). *Spectral approximation of linear operators.* Academic Press, New York.

[20] Chen, X., Hansen, L. P., and Scheinkman, J. (2009). Nonlinear principal components and long-run implications of multivariate diffusions. *The Annals of Statistics*, 37:4279–4312.

[21] Chorowski, J. (2016a). Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies. *http://arxiv.org/abs/1507.07139*.

[22] Chorowski, J. (2016b). Spectral volatility estimators [F# Library]. https://www.mathematik.hu-berlin.de/de/forschung/forschungsgebiete/stochastik/stoch-employees/jakub-chorowski/research.

[23] Chorowski, J. and Trabs, M. (2016). Spectral estimation for diffusions with random sampling times. *Stochastic Processes and their Applications.*

[24] Ciesielski, Z. (1963). Properties of the orthonormal Franklin system. *Studia Mathematica*, XXIII:141–157.

[25] Ciesielski, Z. and Figiel, T. (1982). Spline approximation and Besov spaces on compact manifolds. *Studia Mathematica*, LXXV:13–36.

[26] D. Bakry, I. G. and Ledoux, M. (2014). *Analysis and Geometry of Markov Diffusion Operators.* Springer International Publishing, Switzerland.

[27] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, 19:263–284.

[28] Davies, E. B. (1989). *Heat Kernels and Spectral Theory.* Cambridge Univ. Press.

[29] Davies, E. B. (1995). *Spectral Theory and Differential Operators* . Cambridge Univ. Press.

[30] Dohnal, G. (1987). On Estimating the Diffusion Coefficient. *Journal of Applied Probability*, 24(1):105–114.

[31] Dynkin, E. B. (1965). *Markov Processes.* Springer Berlin Heidelberg, Berlin, Heidelberg.

[32] E. B. Fabes and D. W. Stroock (1986). A new proof of Moser's parabolic Harnack inequality using the old ideas of Nash. *Archive for Rational Mechanics and Analysis*, 96(4):327–338.

[33] Einstein, A. (1905). On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der Physik*, 17:549–560.

[34] Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood Inference for Discretely Observed Nonlinear Diffusions. *Econometrica*, 69(4):959–993.

[35] Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, 20(4):317–357. With comments and a rejoinder by the author.

[36] Fischer, M. and Nappo, G. (2010). On the moments of the modulus of continuity of Itô processes. *Stochastic Analysis and Applications*, 28:103–122.

[37] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4):547–557.

[38] Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30:790–804.

[39] Friedman, A. (1964). *Partial differential equations of parabolic type.* Prentice-Hall Inc., Englewood Cliffs, N.J., 2nd edition.

[40] Ganychenko, I. (2015). Fast L_2-approximation of integral-type functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 2(2):165–171.

[41] Ganychenko, I., Knopova, V., and Kulik, A. (2015). Accuracy of discrete approximation for integral functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 2(4):401–420.

[42] Ganychenko, I. and Kulik, A. (2014). Rates of approximation of non-smooth integral type functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 1(2):117–126.

[43] Gihman, I. I. and Skorohod, A. V. (1972). *Stochastic differential equations.* Springer, Heidelberg.

[44] Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, 32:2223–2253.

[45] Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, 8(S):S85–118.

[46] Hansen, L. P., Scheinkman, J. A., and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, 86:1–32.

[47] Hoffmann, M. (1999a). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79:135–163.

[48] Hoffmann, M. (1999b). Lp Estimation of the Diffusion Coefficient. *Bernoulli*, 5:447–481.

[49] Hoffmann, M. (2001). On estimating the diffusion coefficient parametric versus nonparametric. *Annales de l'Institut Henri Poincaré*, 37:339–372.

[50] Iglehart, D. L. and Whitt, W. (1970). Multiple Channel Queues in Heavy Traffic. *Advances in Applied Probability*, 2:150–177.

[51] Jacobsen, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Δ-optimality. *Scandinavian Journal of Statistics*, 28:123–149.

[52] Jacobsen, M. (2002). Optimality and Small Δ-Optimality of Martingale Estimating Functions. *Bernoulli*, 8(5):643–668.

[53] Jacod, J. (2000). Non-parametric Kernel Estimation of the Coefficient of a Diffusion. *Scandinavian Journal of Statistics*, 27(1):83–96.

[54] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer-Verlag, New York.

[55] Karlin, S. and Taylor, H. (1981). *A second course in stochastic processes*. Academic Press, London and New York.

[56] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5:299–314.

[57] Kingman, J. F. C. (1962). On Queues in Heavy Traffic. *Journal of the Royal Statistical Society*, 24:383–392.

[58] Kohatsu-Higa, A., Makhlouf, A., and Ngo, H. L. (2014). Approximations of non-smooth integral type functionals of one dimensional diffusion processes. *Stochastic Processes and their Applications*, 124:1881 – 1909.

[59] Krugman, P. R. (1991). Target Zones and Exchange Rate Dynamics. *The Quarterly Journal of Economics*, 106:669–682.

[60] Lepski, O. V. (1991). On a problem of adaptive estimation in gaussian white noise. *Theory of Probability & Its Applications*, 35(3):454–466.

[61] Minc, H. (1988). *Nonnegative Matrices*. A Wiley Interscience Publication, New York.

[62] Nakatsukasa, Y. (2010). Absolute and relative Weyl theorems for generalized eigenvalue problem. *Linear Algebra Appl.*, 432:242–248.

[63] Nakatsukasa, Y. (2011). *Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition*. PhD thesis, University of California, Davis, CA, USA. AAI3482268.

[64] Ngo, H.-L. and Ogawa, S. (2011). On the discrete approximation of occupation time of diffusion processes. *Electronic Journal of Statistics*, 5:1374–1393.

[65] Nickl, R. and Söhl, J. (2015). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *arXiv preprint arXiv:1510.05526*.

[66] Papaspiliopoulos, O., Pokern, Y., Roberts, G. O., and Stuart, A. M. (2012). Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531.

[67] Pavliotis, G. (2014). *Stochastic Processes and Applications. Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer-Verlag New York, New York, 1 edition.

[68] Pedersen, A. R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

[69] Pokern, Y., Stuart, A., and van Zanten, J. (2013). Posterior consistency via precision operators for bayesian nonparametric drift estimation in {SDEs}. *Stochastic Processes and their Applications*, 123(2):603 – 628.

[70] Poulsen, R. (1999). Approximate maximum likelihood estimation of discretely observed diffusion processes. *Working paper series / Centre for Analytical Finance, University of Aarhus*.

[71] Qian, Z. and Zheng, W. (2002). Sharp bounds for transition probability densities of a class of diffusions. *C. R. Acad. Sci. Paris*, I 335:953–957.

[72] Reiß, M. (2006). Nonparametric volatility estimation on the real line from low-frequency data. *The Art of Semiparametrics*, pages 32–48.

[73] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, volume I, II. Springer, New York, 3 edition.

[74] Ricciardi, L. M. (1986). Stochastic Population Theory: Diffusion Processes. *Mathematical Ecology Biomathematics*, 17:191–238.

[75] Rozkosz, A. (1992). On Convergence Of Transition Probability Densities Of One-Dimensional Diffusions. *Stochastics and Stochastic Reports*, 40:3-4:195–207.

[76] Rozkosz, A. and Słomiński, L. (1997). On stability and existence of solutions of SDEs with reflection at the boundary. *Stochastic processes and their applications*, 68:285–302.

[77] Shepp, L. A. and Shiryaev, A. N. (1993). A new look at pricing of the "Russian Option". *Theory Probab. Appl.*, 39(1):103–119.

[78] Skorokhod, A. V. (1961). Stochastic Equations for Diffusion Processes in a Bounded Region. *Theory of Probability & Its Applications*, 6(3):264–274.

[79] Skorokhod, A. V. (1962). Stochastic Equations for Diffusion Processes in a Bounded Region. II. *Theory of Probability & Its Applications*, 7(1):3–23.

[80] Słomiński, L. (2001). Euler' s approximations of solutions of SDEs with reflecting boundary. *Stochastic Processes and their Applications*, 94:317–337.

[81] Smoluchowski, M. V. (1916). Über Brownsche Molekularbewegung unter Einwirkung äußerer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffusionsgleichung. *Annalen der Physik*, 353(24):1103–1112.

[82] Söhl, J. and Trabs, M. (2014). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *arXiv preprint arXiv:1412.7103*.

[83] Sørensen, H. (2004). Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: A Survey. *International Statistical Review*, 72:337–354.

[84] Stewart, G. and Sun, J. (1990). *Matrix Perturbation Theory.* Academic Press, Inc.

[85] Stroock, D. W. (1988). Diffusion Semigroups Corresponding to Uniformly Elliptic Divergence Form Operators. *Seminaire de Probabilites*, XXII:316–347.

[86] Stroock, D. W. and Varadhan, S. R. S. (1971). Diffusion processes with boundary conditions. *Communications on Pure and Applied Mathematics*, 24(2):147–225.

[87] Svensson, L. E. O. (1991). The term structure of interest rate differentials in a target zone. *Journal of Monetary Economics*, 28(1):87 – 116.

[88] Tankov, P. (2003). *Financial Modelling with Jump Processes.* Chapman and Hall/CRC Financial Mathematics Series. CRC Press.

[89] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[90] Tuan, P. D. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Series Statistics*, 12(1):61–73.

[91] van der Meulen, F. and van Zanten, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli*, 19(1):44–63.

[92] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

[93] Veretennikov, A. Y. (1979). On the strong solutions of stochastic differential equations. *Theory of Probability and its Applications*, 24:354–366.

[94] Watanabe, S. (1984). *Lectures on stochastic differential equations and Malliavin calculus.* Springer, Berlin, Heidelberg, New York.

[95] Wiener, N. (1923). Differential space. *Journal of Mathematical Physics*, 2:131–174.

[96] Yang, X. and Zhang, T. (2013). Estimates of Heat Kernels with Neumann Boundary Conditions. *Potential Analysis*, 38:549–572.

List of Figures

2.1. Construction by reflection. 22
2.2. Construction by a time change 22

3.1. Different sampling distributions 41
3.2. Volatility functions estimated using the adapted procedure 42

4.1. Comparison of the volatility estimation in high and low-frequency regimes . . 69
4.2. GHR estimator in high-frequency regime 70
4.3. Examples of eigenfunctions in high and low-frequency regimes 71

5.1. Empirical convergence rates of the occupation time 118

List of Tables

3.1. Volatility estimation errors under different sampling schemes 41

4.1. Estimation errors in high and low-frequency regimes 70
4.2. Estimation errors of the GHR estimator in the high-frequency regime 71

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 14 Juni 2016 Jakub Chorowski